

FROM DATA TO JOURNALISM

A BEST PRACTICES GUIDEBOOK

Asad Baig

Sadaf Khan



FROM DATA TO JOURNALISM

A BEST PRACTICES GUIDEBOOK

Authors: Asad Baig and Sadaf Khan

Review: Lina Elter

Design: Nida Meyer Mian

Illustrations: Sana Hammad

FOREWORD

The enhancement of state-citizen dialogue is a key area of the innovative measure Implementing Right to Information Acts – Promoting Data-Driven Journalism implemented by the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and DW Akademie. The project aims for the effective use of the Right to Information (RTI) laws in the provinces of Punjab and Khyber Pakhtunkhwa. Besides streamlining the work of Public Information Officers vis-a-vis the RTI law, the project also focuses on supporting journalists to understand and utilize data-driven journalism methods for effective usage of data acquired through RTI laws.

The media blitz during the last decade or so has changed the very dynamics of communication. The development of digital technology has no doubt facilitated quick transmission of news, but it has simultaneously made us to experience an era of journalism in time of fake news. It is here that the RTI law comes into play, facilitating and ensuring the provision of certified and credible data based information to journalists by public bodies.

By bringing together digital techniques of data analysis and presentation for the enhanced understanding of the news audience and government data acquired through using the RTI laws, this manual will create the space for comprehensive journalism, allowing media to better serve its role in ensuring transparency and promoting accountability. Equipped with these skills and the law, it will thereafter depend on the intellect and the abilities of each journalist to analyze/evaluate this data, and produce something which is worth reading and worthy of appreciation.

**Chief Commissioner
Right to Information Commission
Khyber Pakhtunkhwa**

HOW TO USE THIS GUIDEBOOK

This manual has been designed to introduce users to basic data journalism concepts. It is divided into four sections, the first dealing with the theory behind data journalism to introduce journalists to the basic concepts involved in producing journalistic stories using computer assisted data analysis, the second section focuses on acquiring and managing data and data sets, the third section focuses on skills required to analyse data sets and the fourth deals with actually telling stories with this data.

For the ease of users' the manual uses just a single dataset in all the exercises that have been included for practice. The data set is available in the USB given with this manual or can be downloaded from: <http://mediamatters.pk/wp-content/uploads/2017/05/Casualties-of-Media-Workers.xlsx>



TABLE OF CONTENTS

SECTION 1

Introduction to data journalism

Chapter 1 – What is data journalism?	10
Chapter 2 – The data journalism process	14
Chapter 3 – Becoming a data journalist	18

SECTION 2

Finding and understanding data

Chapter 4 – The data hunt	22
Chapter 5 – RTI laws and applications	29
Chapter 6 – The quantification of information	46

SECTION 3

Making sense of data

Chapter 7 – Interviewing the data set	58
Chapter 8 – Spreadsheets, the data journalists' secret weapon	63

SECTION 4

For the eye of the beholder

Chapter 9 – Basics of data visualization	82
Chapter 10 – Data mapping	94

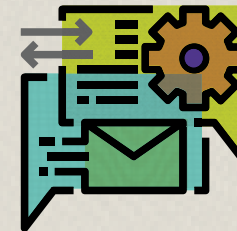
SECTION 5

Additional Resources

Chapter 11 – An ongoing quest	106
-------------------------------	-----

ANNEXURES

114



“IN GOD WE TRUST, ALL OTHERS
BRING DATA,”

- W. Edwards Deming
Physicist and Statistician

SECTION
ONE

INTRODUCTION TO DATA JOURNALISM



CHAPTER ONE

WHAT IS DATA JOURNALISM?

IMAGINE being able to prove beyond doubt the exact nature and volume of transactions that were made to create the offshore companies revealed in Panama Papers. Imagine, being able to demonstrate the volume and impact of the circular debt on power pricing and production. The numbers you will need to make these demonstrations constitute the ‘data’ in data journalism.

Defining data journalism

Data journalism is simply a form of journalism in which numbers are used to tell stories in an effective and demonstrative manner. The term ‘data’ often intimidates journalists by bringing back memories of difficult mathematics and statistics problems that they might have struggled with. That’s why it is important to make it clear in the beginning that while data journalism does include numbers, it is by no means just about mathematics or creating charts or graphics, but is actually about utilizing those numbers for telling interesting stories. So, the numbers or the data simply adds to the arsenal that you, as a reporter, generally use to tell the story (like words, audio, photographs or video). Instead of documentary or visual ‘proof’, data journalists use authentic numerical data as a way to substantiate the story that they are telling.

So, the process of data journalism essentially includes using numbers as a tool to support the story you are going to tell.

What is data journalism?

The British Broadcasting Company, BBC, states that data Journalism “covers projects that use data to do one or more of the following:

1. Enable a reader to discover information that is personally relevant
2. Reveal a story that is remarkable and previously unknown
3. Help the reader to better understand a complex issue.

As the BBC definition demonstrates, data journalism is also about finding stories within numbers. The data isn’t only a tool for telling a good story as mentioned earlier, but also an important part of the story itself.

While it isn’t necessary to be able to perform complex mathematical functions with the data you are working with, it is however, very important to understand the story that the data is telling. Once you understand what the numbers represent, you can use them effectively within your journalistic stories.

“Data journalism is the use of data and number crunching in journalism to uncover, better explain and/or provide context to a news story”
- Technopedia

Defining data

It is difficult to coin an exact definition of what constitutes data in general terms. Within the data journalism context however, data has to be something measurable or quantifiable. The definition of data by Poynter Institute, one of the global leaders in academic research on data journalism lays a broadly defined field by terming everything countable as data. This means that for creating data based stories, you need to look at countable elements present within the story’s premise. For example; for a story looking at the state of the healthcare system in a country, the countable elements can be doctor to patient ratio in different cities, the patient to bed ratio in a certain hospital, the allotments of health budget in different quarters or in successive governments.

“Everything countable can count as data. Similarly, everything that a computer can compute is also data. So vast is the actual definition of data, that it is impossible to limit what data journalism actually constitutes”
- Poynter Institute

Each of the elements mentioned in the example i.e. the number of doctors, patients, beds, money is quantifiable and countable.

So, while the larger story might include interviews and sound bites from people who are trying to access the healthcare system, the

data based elements of the story would be concerned with the numbers linked to that story. In most cases, these numbers add different dimensions to the whole story. For example, the numbers looking at doctor to patient ratio in different cities might demonstrate trends about the difference in affluent and peripheral cities or towns.

You might be wondering the sudden need to focus on data journalism, when the example given above simply talks about using data in a way that you and your fellow journalists are already familiar with. After all, whenever official figures are available, they become a part of your stories. So, question is, **what exactly is new about data journalism?**

The new thing about data today

There are three basic factors that make data journalism so relevant today. First, the number of tools available to analyze data has increased exponentially. In the past, a reporter in possession of a vast set of documents that included a lot of numbers would have to manually dredge his or her way through them to find out what the numbers revealed. At the most, the one tool available to him or her would be a calculator. In essence then, he or she would be mostly relying on the analysis that was already done by other researchers or statisticians. Despite this barrier, numbers were always used by media organizations, particularly for finance related reporting. Today’s reporters however have a vast number of easy to use, easy to access tools that allow them to filter, manage, analyze, understand and visualize data. Be it your smart phone or your computer, both are equipped with multiple tools that can aid and

assist you in data journalism. This is why data journalism is at times also referred to as **computer assisted journalism**.

The second factor aiding the practice of data journalism is open data. According to the Open Data Handbook¹, the term open data is used to define data sets that "can be freely used, reused and redistributed by anyone – subject at most to the requirement to attribute and share alike". Open data is fast becoming a part of the development and governance jargon. Pakistan became a part of the Open Government Partnership in February 2017 – which means that the government now is dedicated to making serious efforts to open up its data for transparency and accountability.

The term open data itself is connected to open governance and entered the mainstream in 2009 after governments in Canada, UK, USA and New Zealand opened up large sets of government data in useable formats. The utility of open data is uncontested and its economic, political and social benefits are being documented. In addition to the governments, multilateral bodies including various agencies of the United Nations also release large data sets in open formats that can serve as a rich resource to journalists.

The third factor that makes data journalism so relevant today is the changing face of journalism itself.

REQUISITES OF OPEN DATA

1. **Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
2. **Re-use and Redistribution:** the data must be provided under terms that permit re-use and redistribution including the inter-mixing with other datasets.
3. **Universal Participation:** everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavor or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

- The Open Data Handbook

As media companies move towards convergence of mediums and the digital devices become one of the main platforms of news consumption, data visualizations and interactive data based infographics are fast becoming a very important visual element of news stories. The trends in the global news market and economy demonstrate that the trend of mobile first as a platform for news distribution will eventually arrive in Pakistan as well. Thus, it is important for journalists of today to acquaint themselves with basic tools and acquire the necessary skills to conceive, manage and design data journalism based projects.

THINGS TO REMEMBER

1. Data journalism is simply the use of data, mostly through computer-assisted analysis to either find or support a story.
2. Data, in the context of data journalism has to be measurable, countable and quantifiable.
3. Data journalism isn't a new phenomenon, but developments in technology, digital tools for data analysis and visualization, presence of open data and the changing dynamics of news markets are making it more and more relevant in today's journalism industry.
4. With Pakistan becoming a member of the Open Government Partnership and the Right to Information laws taking shape, the opportunities and potential for local data based stories has increased exponentially.



CHAPTER TWO

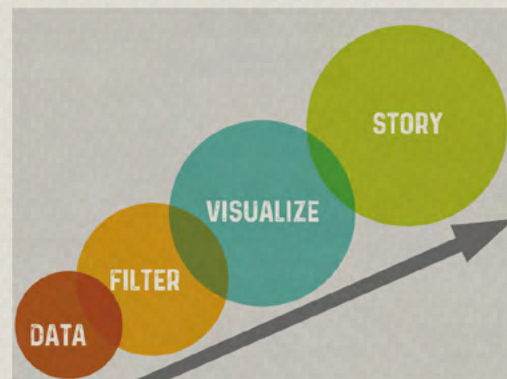
THE DATA JOURNALISM PROCESS

THE data journalism process is much like the one you are already employing at some level to create journalistic stories. It deals with processing information, to identify something newsworthy and then presenting it in a publicly consumable format as a data based news story. In this chapter we will take a brief look at the different steps involved in the data journalism process. However, it is important to note that the process is fluid and in different instances, different parts of the process can occupy different positions.

The descriptions of the data journalism process by known journalists with data expertise highlight a similar process. Paul Bradshaw, a data journalists who has worked extensively with the Guardian Data Blog describes the process of data journalism by saying that first, data must be found, which may require specialized skills like MySQL or Python, then interrogated, for which understanding of jargon and statistics is necessary, and finally visualized and mashed with the aid of open source tools.²

The process defined here is similar to the one used by Mirko Lorenz, a data and multimedia journalist, who sees data-driven journalism primarily as a workflow that consists of digging deep into data by scraping, cleansing and structuring it, filtering by mining for specific information, visualizing and making a story.³

Here is a visual representation of the data journalism process



The data journalism process visualized by Mirko Lorenz, sourced through Wikipedia

2. Read Bradshaw's full article here - <https://www.theguardian.com/news/datablog/2010/oct/01/data-journalism-how-to-guide>
3. See an analysis of the data journalism process by Lorenz here - <http://www.slideshare.net/mirkolorenz/datadriven-journalism-what-is-there-to-learn>

Data

The first step of the process is the acquisition of data. Here think back to the definition of data we discussed in Chapter One. The data we have to acquire for data driven journalistic stories has to be in a countable format. Now, as a journalist you are in possession of a lot of information. For example, you might be getting regular updates from your local police station about the registered crimes in different localities. Or if you cover the assemblies or the parliament, you have information about the participation of different Members of National Assembly, MNAs and Members of Provincial Assembly, MPAs during assembly sessions. However, in its raw form, this information might be descriptive and textual. It might be abstract. For you to write a data based story about it, you

need to find a way to convert this information into a quantifiable or countable format. So, the acquisition of data, might either mean gaining access to open data sets, other kinds of available data or it might mean sifting through the descriptive information you have to find a way to convert it into data sets. In the following sections, we will go into more detail about the sources through which data can be acquired and managed.

Filter

Once you have access to the data, the second step is to filter it or to analyze it to find the actual story. Take a look at this data set:

Annual Cellular Subscribers						
Year	Jazz	Ufone	Zong	Telenor	Warid	Total
Jazz	33,378,161	20,533,787	10,927,693	26,667,079	17,387,798	108,894,518
Ufone	35,953,434	23,897,261	16,836,983	29,963,722	13,499,835	120,151,235
Zong	37,121,871	24,547,986	21,177,156	32,183,920	12,706,353	127,737,286
Telenor	38,768,346	24,352,717	27,197,048	36,571,820	13,084,823	139,974,754
Warid	33,424,268	17,809,315	22,102,968	31,491,263	9,830,620	114,658,434
Total	39,118,521	19,833,670	25,251,329	38,020,771	11,017,174	133,241,465

Source: Pakistan Telecommunication Authority, PTA

The process of filtering and analyzing this data would mean looking at the different variables present to find out what this data is actually telling you, and then figuring out what exactly you want your story to be about. So, in this data set your story could be about the comparative growth of different cellular companies during six years. Or you can choose to ignore the company wise data and simply look at the total, highlighting the cellular usage in the country. At times, at this stage of the data journalism process, you will also be looking at other related data sets.

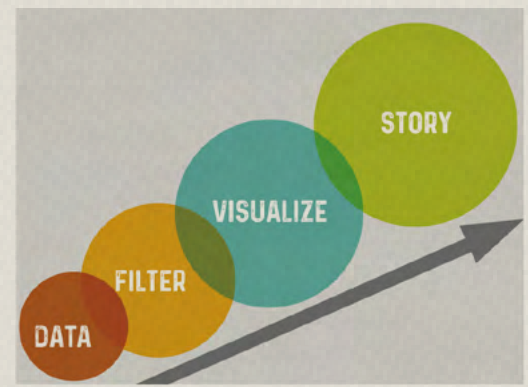
Visualize

The first two steps of the data journalism process are more related to news production. The visualization process is both for you, the journalist working on the story and for your reader. Putting the data into visualized formats be it graphs, tables, maps or infographics will first help you understand your story better and more easily. And more importantly, your readers, who might be intimidated or confused by large sets of numbers would be able to visually make sense of the story without having to put an additional effort to understand what the numbers actually mean. Visualization process involves finding the right tools and the correct type of visualization that best tells the story.

If we reproduce the data by PTA as a table, it will be difficult for a general news consumer to understand. One of the most common ways to visualize data of this kind is through graphs. So, if your story is about the comparative growth of different cellular companies, you might want to visualize this through the creation of a trend line. Or we might create an area graph to demonstrate what percentage of the market each company has. In other cases, the data might be plotted on a map or at times, it might be so complex that you would need to develop an infographic including multiple sets of visualizations to properly demonstrate the story to your reader. The basics of visualization, including the considerations you have to make while selecting the most effective visual tool for your reader, are discussed in Section Four of this guidebook.

Story

The fourth part of the data journalism process has simply been termed as story. This means putting together the necessary context to make sure that your reader understands the story you are trying to tell with the data. The story telling part of the process also includes telling the reader why is this important. For example, what is the impact of the growing number of cell phone subscribers, what does it demonstrate about the economy or accessibility. Your story might pull in other related data sets and their analysis. In the PTA data, you might look at the policy changes that have occurred over the years and your story might be about the impact of these policies over the growth of the cellular industry. In a nutshell, the story telling part of the process means giving life, depth, context and value to the effort that you have made in acquiring, analyzing and visualizing the data. It is this part of the process which will hook your reader and make him/her understand why you find it important to make public your analysis of the data and what news value does your data visualization has.



If you look at the visualization of the data journalism process again, you will also notice an arrow that says ‘rising value to public’. What this means is, as you move through the different steps of data journalism process, the product you have is of increasing value for public or public consumption. Going back to the example we have taken in this chapter, even as the table, the data you acquired from PTA has an informative value to the public. However, once you filter and analyze it, the value increases as the data now converts from simple information to information that demonstrates a certain story. Third, when you visualize this filtered and analyzed data, the story that you are telling with it becomes easier to understand for the reader, thus the value of the data increases further. And finally, when you create your data story, you add the background, details, context or relation to other data/ indicators etc. which adds both depth and beauty to the product and its value goes further up. The Verification Handbook for Investigative Reporting visualizes the process of data journalism or data based investigations in a cyclic manner.

Phases of an investigation with data



Source: Verification Handbook for Investigative Reporting

This cycle does not include the story telling step of the data journalism process we discussed earlier. The story can only be started once this cycle of investigation with the data is complete, which means that at times after visualizing the data you might realize a new angle, a new aspect of the story that you had previously missed and find yet more data to add new dimensions.

THINGS TO REMEMBER

The data journalism process includes four steps

1. Acquisition of Data – the source of your story or the tool with which you are telling the story
2. Data filtration & analysis – using digital tools to analyze the data and give it meaning
3. Data Visualization – Using visual tools to display the trends or facts that your data demonstrates
4. Story telling – adding the background and other details to contextualize your story

CHAPTER THREE

BECOMING A DATA JOURNALIST

NOW that you understand what data journalism means and the process through which data driven stories are created, we can look at the skills, tools and techniques that you require to become a data journalist and start creating data based news stories.

In this chapter, we will introduce you to the different tools that we will be using throughout this guidebook. These are the basic tools and techniques that you need to go through the data journalism process. But before we list the tools that we will be using, it is important to note that the tools and techniques that are required for data investigation are forever evolving. As technology develops, newer tools for data analysis and visualizations become available. Keeping track of these tools will help you add new techniques to your repository and create more interesting stories.

So, what do you need to become a data journalist?

A simple answer would be “data” – but then the data itself is never enough. What a data journalist needs is also the context, the question that the data can answer. So, when you start a data based project, it is always good to start with a beat that you already understand. If you are a crime reporter, find a story within the crime beat that is best told by using numbers and drawing trends. If you are covering the health beat, again, what are the stories that you can tell best if you had the data to demonstrate what you have found. The first thing you need then is **an eye for data and story angles that benefit from numbers**. To develop an eye for data, you need to know the main sources of data in your beat and subscribe to those sources. To understand story angles that depend on data based investigations, you can keep a track of data stories being written at international platforms to get inspiration. A skilled data journalist should also be proficient in contextualizing the initial data set with other related datasets.



First, a computer of course!

1. An eye for data
2. The knack to find data related angles in an issue
3. A tool to store, manage and analyze the data
4. Conceptual understanding of data visualization
5. Technical skill to create the visualizations
6. The will to put all his or her effort into creating data based stories

Once you have acquired the data you will need the tools to clean, filter and analyze it. In this guidebook, we will introduce you to the most basic tools for performing these functions; Google Forms for data management and Microsoft Excel for data analysis. There are other tools that are available either for free use or for a minimal cost including Google Refine, Scrapper Wiki etc. that will allow you to manage data sets in workable formats. These tools are slightly more advanced than the skill set being targeted in this guidebook.

Once you have the data in a workable format and are able to analyze it, you will need the skill to visualize the story your data demonstrates.

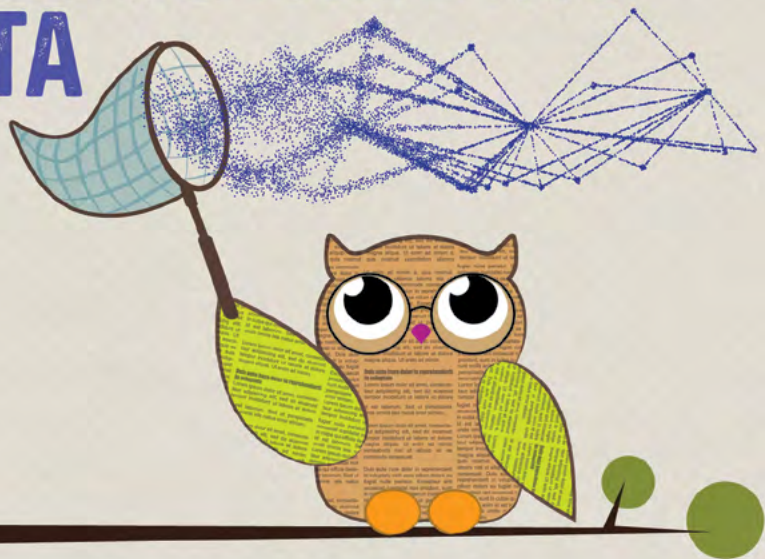
Data visualization needs both conceptual and technical skills. Conceptually, you have to be able to understand which parts of the data should be included in the visualization and which kind of visualization is most effective for understanding the data. Technically, you should be able to create tables, graphs, charts, infographics and maps. In this guidebook, we will focus on basics of data visualization, that will help you understand the considerations you have to make when visualizing a data set. The technical tools we will introduce you to include **Infogram** and **Google Fusion Tables**.

THINGS TO REMEMBER

1. Data journalism requires both conceptual and technical skills.
2. Neither of these skills are difficult to achieve!
3. You also need an open mind to explore all the possible stories that a dataset has to offer.
4. Data driven journalism is computer assisted so, you will need access to a laptop or a desktop
5. The basic tools required for technical aspects of data based investigations are either free or quite cheap and are user friendly.

FINDING AND UNDERSTANDING DATA

SECTION TWO



CHAPTER FOUR

THE DATA HUNT

FOR most of the Pakistani journalists ‘data’ usually represents an elusive treat. In our interactions with journalists in different cities and different organizations, the availability of data has always been presented as the first and the foremost matter of concern. In this chapter, we will take a look at some of the sources of open data that can be utilized by journalists for data acquisition. We will also discuss how to make best use of these sources.

Identification of data sources

Let us explore the process of data hunting by taking the example of parliamentary and legislative data. First, most government bodies have some data that they publish online. Exploring websites registered in the dot gov dot pk (.gov.pk) domain will give you access to a lot of parliamentary data. This data might not be in open data formats but it is still pretty easily accessible and workable. The National Assembly’s website has a wealth of information.

This screenshot of the National Assemblies’ website shows the wealth of data available online. This data includes all debates that have occurred on the floor of the house since 1947, the first session of the constituent assembly. In the members’ section, you can also access the attendance details of all MNA’s. Some of the documents include rich information that needs some effort before it can be transformed into data. Regardless, in a country whose media thrives on politics, this is a rich source of data that is not used as often as it could be. The websites of provincial assemblies also have a lot of information and data easily available, even though none of them are quite as detailed



as the national assembly. Reporters covering the assemblies can also get further information and data from the assemblies’ libraries.

If it is political data you are after, it is also important to keep track of other organisations that monitor and report on government’s performance. For example Pakistan Institute of Legislative Development and Transparency, PILDAT regularly monitors and comments on the performance of parliamentarians. PILDAT’s reports shared through its website <http://www.pildat.org> can be another good source of data related to parliamentarians. Another source of data about parliamentarians is Election Commission of Pakistan, ECP.

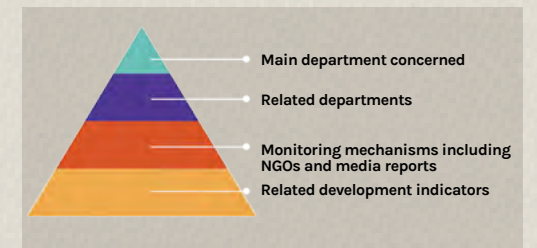
The ECP can be contacted for acquisition of details of parliamentarian’s assets and liabilities. Similarly, the Federal Bureau of Revenue, FBR, can be contacted for the acquisition of data about tax returns filed by parliamentarians. The Pakistan Institute of Parliamentary Services, PIPS, can be contacted for acquisition of data regarding the legislative support requested by parliamentarians.

What we have done in this example is go through a process of mapping some institutions that relate to the beat in question i.e. legislative and parliamentary proceedings. The process of data hunting includes;

1. Identification of the key department [institution or organization] that is linked to the story you are exploring.
2. Identification of related bodies – these might be official institutions, or ministry’s whose policies guide the working of the institution in focus or non-governmental bodies including NGOs that monitor the body in focus.
3. Exploration of available open data – data that is available digitally on websites.
4. Requisition of data not available online – the requisition can be done through a source or through the legal right to information mechanisms that are present.

Now related bodies that connect with cases of violence against women can be

- i. National / Provincial Commissions on Status of Women
- ii. Human Rights Ministry
- iii. National Commission for Human Rights
- iv. Women Parliamentarians
- v. Dar ul Aman



Possible data sources when exploring a data story

The monitoring bodies and mechanisms can include

- i. Aurat Foundation
- ii. Shirkatgah
- iii. Women Action Form
- iv. Media reports

The related development indicators can be

- i. Databases maintained by UN Women
- ii. Gender Gap reports
- iii. Gender Equality Index

What we are doing here is simply brainstorming to see which data sources can be picked to add some value to our story.



Exercise

You are working on a story about the spread of dengue fever in your province / region;

- i. Which department is your main source of data?
- ii. Which are the other official bodies that can provide you with related information?
- iii. Which are the non governmental bodies or entities engaged in monitoring that can provide additional information?
- iv. Are there any related global indicators that can be checked for context?

Google for Journalists

Like all journalists of the digital age, you must already be familiar with Google. The search engine, that is almost synonymous with the internet and a brand name that is used as a verb has several features that can be of extraordinary help to journalists. Recognizing this, Google itself hosts a repository of tutorials for journalists at its initiative Google NewsLabs. Accessible via <https://newslab.withgoogle.com> the site contains various tutorials to help journalists increase the efficiency of their searches, use Google for verification and for production and dissemination of news.



Google NewsLab – a virtual training platform

At the moment, Google NewsLabs hosts tutorials about tools and applications linked to trust and verification, data journalism, immersive storytelling and inclusive storytelling. In this chapter, we will be looking at Google Advanced search options.

Imagine yourself googling for something that Mustafa Kamal said while he was a part of the Muttahida Qaumi Movement, MQM. Given his new political affiliation, it might be very difficult for you to wade back through all the information to find his older statements.

From the horses' mouth

You can find out more about the advanced search features in a YouTube video by Google accessible at <https://youtu.be/FtSeAwRhPKs>

If only you could search MQM's own archives for statements attributed to him or just skip right past the more recent statements. Google advanced search options will allow you to do all this and more.

Below we explain the basic advanced search features.

1. Eliminating unwanted search results

Or as Google puts it, search for this. Not that. This search feature allows you to create a basic filter within your search query. To do this, simply use the '-' key before the word that you want eliminated from the query. So, if you want to search for Mustafa Kamal but not Pak Sar Zameen Party, PSP, you can simply add -PSP at the end of your query. This will eliminate results where PSP is mentioned. Remember, there should be no space between the '-' symbol and the word that you want to eliminate from the results.

2. Searching fro a specific site

Let us say that you are searching for a story about Panama Papers that you read in The News. Now simply putting Panama Papers in the Google search bar will give you innumerable results. To make the search more effective, you can use the modifier 'site:URL' to limit your search to a particular site. In this case, your search query would be Panama Papers site: <https://www.thenews.com.pk>. Again, there should not be any space between site: and the URL.

3. Searching for a particular filetype

If you are looking for data that is already clean and in reusable, open format, Google allows you to specify that within your search query. Some of you might already be doing that – for example when you add .pdf to your search query in an attempt to find books, or research articles. The modifier you can use to limit the searches only to pdfs, or xls or doc or other specific formats is filetype: formatname for example, filetype: .xls will bring search results with spreadsheets.

What is search query?

The phrase search query, in this guidebook refers to the combination of key words and modifiers that are being used in the search bar of search engines like Google.

4. Getting multiple search terms

It is also possible to create a search query that allows you to get multiple key words. For example, if you want to want to watch the infamous Aam khaye ga aam video you can use the '+' sign to search for **Amir+Liaqat+Aam**. This will make sure that you get results where all three key words appear.

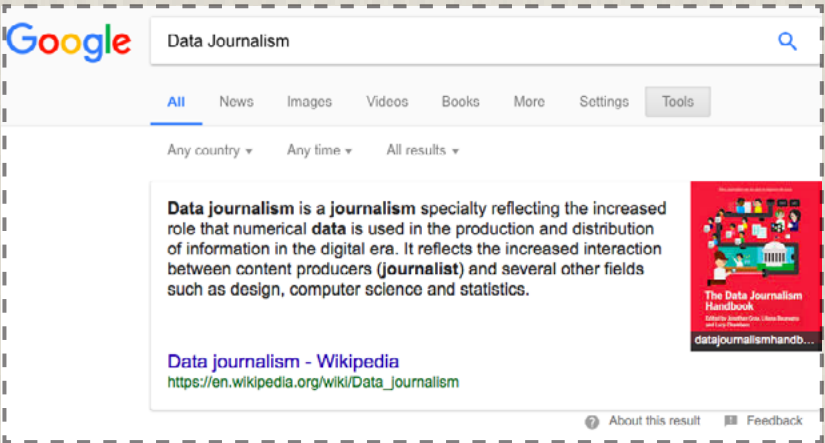
Advanced search via Google

If you find the modifiers difficult to use or remember, you can also take advantage of the built in advanced search feature. To do that go to google.com, add any test search query.

Sample screenshot of a search result in Google



On this search result, you can see the **settings** and **tools** tabs.

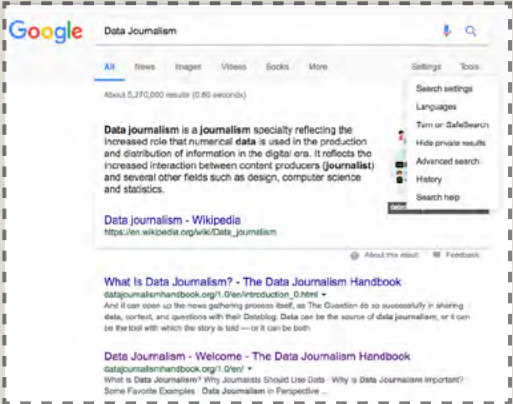


Expanding the Tools Tab

On clicking the Tools tab, you will get three options, among which these two are particularly helpful.

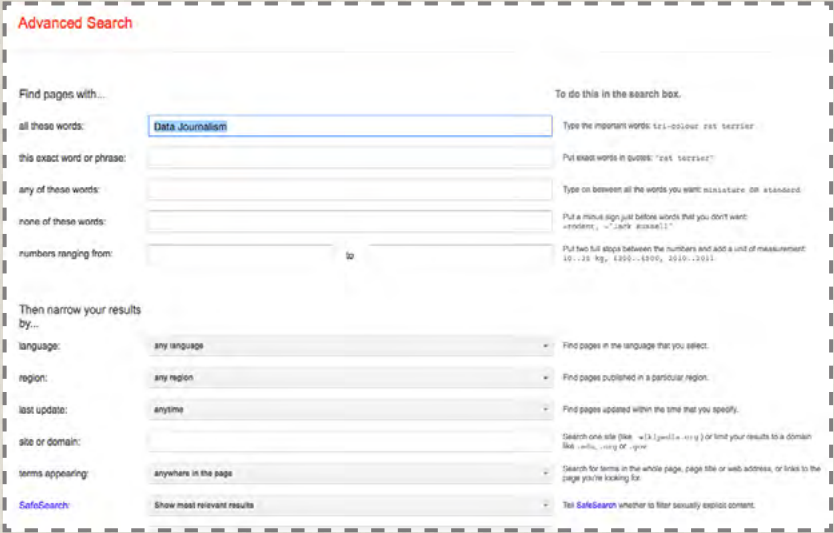
- I. **Any Country** – that allows you to select the country from where you want your search results. For example, people who are interested in history might try searching for accounts of the 1971 war from websites in Bangladesh to get a different perspective.
- II. **Any Time** – this tab allows you to choose the duration during which the item, or information that you are looking for should have been uploaded. You can choose to get results from the last 24 hours, the last week, last month or year or anytime.

There is another option on this page, that will allow you to create even more focused and efficient search queries. That option is the advanced search option, which you can find by clicking the Settings tab.



Sample screenshot of a search result in Google

The advanced search option allows you to automate all the modifications that you learnt earlier in this chapter.



The advanced search page

The top section titled Find pages with, of the advanced search page deals with the modifiers we already discussed earlier in this section. By adding key words in the different fields of this section you can find results with

- Multiple key words
- Words in a specific sequence
- Searches eliminating certain words
- And / or search results containing certain numbers.

Additionally, the narrow your search options allow you to

- Select your preferred language
- The region where you want your search results to be from
- The time of the last update made on the site
- Search within specific sites
- Choose where the search terms should be positioned within the webpage or website,
- The file type you are looking for
- And look for content with specific usage rights.

The usage rights refer to the nature of copyright license that any piece of content has. This is important if you are looking for images or other media like music to reuse in your own work.

THINGS TO REMEMBER

1. You can use multiple modifiers together. For example, you can eliminate or add search terms and look through a specific site at the same time. Eliminate multiple words, add various words together and find a specific file type while doing that.
2. If you are manually using modifiers, remember adding space after the : sign will corrupt the sequence and you will not find the results you are looking for.
3. As all technological systems, Google works on parameters that can be affected by how an element itself is marked. Which means that if a website has added wrong tags to describe its content, it would appear or not appear in your search accordingly.

CHAPTER FIVE

RIGHT TO INFORMATION LAWS

RIGHT to information, RTI laws have been used world over by journalists to acquire official data from government departments. In Pakistan, the RTI laws have been historically weak, but in 2014 the situation changed for better when Khyber Pakhtunkhwa, KP and Punjab, consecutively passed effective RTI laws. The implementation of these laws has not been perfect and similar laws are still needed in Sindh, Balochistan and the Federal constituencies, but, the situation is still much better than before. In this chapter we will take a look at the RTI laws you can use to gain access to data and the process of making RTI applications.

Right to Information Laws

Let us start with a bit of a context. Right to information is globally recognized as one of the fundamental rights. The United Nations Sustainable Development Goals, SDGs 16.10 aims to **"ensure public access to information and protect fundamental freedoms, in accordance with national legislation and international agreements."** In Pakistan, the right of access to information was granted by a constitutional amendment and is granted to the citizens of Pakistan through Constitutional Article 19A, that states that **"Every citizen shall have the right to have access to information in all matters of public importance subject to regulation and reasonable restrictions imposed by law."**

Thus the right to information has to seen and practiced as both a universal and constitutional right. However, the laws governing this right in Pakistan differ from region to region.

RTI Law in the Federal Jurisdiction

The RTI Law passed by the Parliament of Pakistan applies to the information that is owned by those ministries and departments that are operating at the federal level. The Freedom of Information Bill was originally drafted in 1996, by Federal Law Minister Fakhruddin G. Ibrahim. However he was a part of the interim government and the successive government of Pakistan Muslim League, Nawaz PML N did not enact it into law, causing it to lapse. In 2001, the military government led by Gen. Pervez Musharraf promulgated the Freedom of Information Ordinance 2002 and added certain provisions pertaining to the right of access to information held by district public bodies in the Local Government Ordinance 2001. Both these measures were taken as a result of a conditional loan that was to be attained from Asian Development Bank, ADB.

The RTI law at the federal level is thus a weak law. Various attempts to amend the law have been made by certain parliamentarians in successive governments but so far, the amendment bill has not been passed.

Freedom of Information Ordinance 2002 mandates public bodies to facilitate and encourage, promptly and at the lowest reasonable cost, the disclosure of information (Section 2(ii) FOI Ordinance 2002).

The bill also has provisions for maintenance and indexing of records, publication and availability of records, computerization of records, declaration of public records, that include transactions involving acquisition and disposal of property and expenditure undertaken by a public body in the performance of its duties. Information regarding grant of licenses, allotments and other benefits and privileges and contract and agreements made by a public body, final orders and decisions, including decisions relating to members of public and any other record which may be notified by the Federal Government as public record also falls under the ambit of this law and can be requested.

The ordinance also bars access to a subjective and long list of information. One of the weakest points of this law is the lack of consequences for public officials who fail to comply by the spirit of this law. The Wa-faqi Mohtasib or the Federal Tax Ombudsman can be approached by aggrieved applicants but traditionally their decisions have not been considered binding by the government bodies.

The Information available under FOI Ordinance 2002

Information you can access		Information that you cannot access
1.	Transactions about acquisition and disposal of property	Meeting minutes of public bodies, files and recommendations made before final decisions
2.	The expenditure by public bodies	Banking and financial records of costumers
3.	Information dealing with the grant of license and other contracts and agreements by public bodies	Any information dealing with defense forces and linked to national security

4.	Decisions taken by public bodies	Any record that the government declares classified or the government wants to exclude for public interest
5.	Other records that the federal government chooses to declare public for the sake of this law	Records that deal with information of personal privacy of an individual
6.		Information that might affect Pakistan's relations with other states
7.		Information that can result in offences, effect investigations, reveal identity of a confidential source of information, facilitate escape from custody, harm security of any system
8.		Information that can cause economic loss or damage

Right to Information Law in Khyber Pakhtunkhwa

Khyber Pakhtunkhwa Right to Information Act 2013 is largely considered a success story. The bill is progressive and incorporates the real spirit of the right to information. It also enacts the KP RTI Commission, which is an empowered body where members of the public who feel aggrieved by the responses of government bodies approached for information can file appeals for review and relief.

There are some issues with the implementation of the law but by and large, civil society organizations working on RTI appear happy with the content of the bill and there are efforts to make sure that the implementation improves over time. The RTI Commission in KP has also been active and is also involved in encouraging departments to predisclose information in line with the spirit of the law.

Which bodies are liable to provide information under KP RTI Act 2013

- All public bodies being paid for through the provincial exchequer are obliged to provide citizens with information, including:
- All government departments (e.g. Department of Education, Local Government Department)
- All attached organizations (e.g. Highways Department, KPK Police, Directorate General Public Relations)
- All autonomous bodies (e.g. KMC College, KPK Auqaf Organization, Boards of Intermediate and Secondary Education, PDA, Universities etc.)
- All local governments (e.g. your Tehsil Municipal Administration)
- Any statutory body set up under a provincial law (e.g. the KPK Ombudsman Office, KPK Public Service Commission)
- Subordinate Judiciary
- The KPK Assembly (e.g. your MPA)
- The Secretariats of the Governor and Chief Minister of KPK
- Any Board, Commission, Council or a body financed by the Government or
- Any other body which undertakes public functions.

- Via KP Information Commission

Not only the Act governs almost all public functionaries, but it also allows all citizens of Pakistan and companies and organizations registered in Pakistan to request public information. Here is a closer look at the information you can and cannot request under the KP RTI Act 2013.

Information available under Khyber Pakhtunkhwa RTI Act 2013⁵

Information you can access		Information that you cannot access
1.	Laws, rules and regulations by which a public body operates	Information that can be proven to impact National defence/ security/ public order; Legally privileged information (e.g. lawyer-client discussions)
2.	Budget both allocated funds and actual expenditure of public bodies	Information that can harm international relations of Pakistan
3.	Organizational set up of public bodies including staff numbers, grades and salaries and seniority lists	Information that compromises someone's privacy like bank account details
4.	Tenders and contract awards; services provided; annual and other reports All other information dealing with the working and decisions of the public body unless specifically exempted	Information that can harm commercial interests
5.		Information that can compromise someone's health or safety for example identifying details of a person charged with blasphemy
6.		Information that can harm administration of justice or tackling crime, like operational details of an ongoing police operation
7.		Information that might harm the government's ability of the Government to manage the economy

Two factors give KP RTI Act 2013 additional strength.

5. Via KP Information Commission

First, none of the exemptions are absolute. The government or the public body that denies any information, even if it is exempt has to demonstrate why they think that disclosing the information will cause harm. Even then, the applicant can challenge the argument in the RTI Commission, which is empowered and whose decision can only be challenged in High Court.

Secondly, if the commission feels that despite the potential of harm, disclosure of certain information will have a larger public interest value, it can direct the public body to disclose information.

Punjab Transparency and Right to Information Act 2013

This is another provincial law that is rated highly on global indicators of RTI laws . Like the KP RTI Law, the implementation of this Act has been challenging. Despite those challenges, the adoption of the law is slowly picking up. Like KP, the law in Punjab also created an RTI Commission, that can be approached by members of public and journalists who feel that they are not getting satisfactory responses from government departments or public bodies approached for information requests. This Act scored

over 123 points, out of maximum 150 points in a set of indicators developed by The Centre for Law and Democracy (CLD), Canada and Access Info Europe (AIE) to measure the strength of RTI Laws.

The law extends to all public bodies in Punjab and bodies that receive public funding or have a definite public function.

Here is a more detailed look at what you can and cannot request under Punjab Transparency and Right to Information Act 2013.

6. A national rating index comparing the different laws on RTI has been developed by Centre for Peace and Development Initiatives, CPDI and is accessible through <http://www.crti.org.pk/wp-content/uploads/2015/11/CPDI-Score-Sheet-of-Right-to-Information-Laws-in-Pakistan.pdf>

Information available under Punjab Transparency and Right to Information Act 2013⁷

Information you can access		Information that you cannot access
1.	The laws, rules, by-laws and regulations by which a public body operates	Information that can be proven to impact National security
2.	Information regarding the budget and expenditure of a public body	Information that can harm the economy
3.	Information regarding the announcement and procedure for tenders and contract awards	Information that compromises someone's privacy
4.	Information about services provided, process and official fees (if any) for obtaining those services	
5.	Annual and other reports of public bodies	
6.	Any other information related to public bodies that is not specifically exempted under Section 13 of the Act	

The Act allows any citizen of Pakistan to ask for any information in any type of information like memo, book, design, map, contract, representation, order notification, plan, accounts statement, project proposal, photograph, audio, video, any instrument prepared through electronic process, or any machine-readable document.⁸ The information can also be accessed in physical form; even samples of construction materials can be requested.

Like the KP Act, this law also states that no exception is absolute and the Commission can order access to exempt information if it determines that the benefits of disclosure outweighs the potential harm.

8. Via Punjab Information Commission

RTI Law in Sindh

The Sindh Assembly has just passed a new and improved version of provincial RTI Law on 13th March 217. According to media reports” the legislation envisages a proper mechanism with defined timelines for citizens to obtain information on any public authority in the province as well as proposes a system of penal action against officials withholding or obstructing release of information.”

Public body includes departments, attached departments, autonomous and semiautonomous body of government, companies and special institutes created by the government. local government constituted under sign local government Act or any other law, secretariat of Government of Sindh. any court, tribunal, board, commission, council or others body functioning in Sindh. Sindh Assembly, statutory bodies established under provincial laws and NGOs getting substantial funds by government and secretariat of chief minister Sindh.

Like the Punjab Act, this bill also extends the definition of information declaring information” as any datum held by a public body and includes any memo, book, design, contract, representation, pamphlet, brochure, order, notification, document, plan, letter, report, accounts statement, project proposal, photograph, audio, video, drawing, film, electronically prepared instrument, ma-

chine-readable document and other documentary material, regardless of its physical form or characteristics.” (CLD), Canada and Access Info Europe (AIE) to measure the strength of RTI Laws.

The law extends to all public bodies in Punjab and bodies that receive public funding or have a definite public function.

Here is a more detailed look at what you can and cannot request under Punjab Transparency and Right to Information Act 2013.

RTI Law in Balochistan

The Balochistan Freedom of Information Act 2005 is an exact replica of Freedom of Information Ordinance 2002 and is seen as a very weak and toothless piece of legislation. Advocacy for the amendment of this bill has continued over the last decade without much success.

Requesting Information Under RTI Laws					
	Cost ⁹	Requirement ¹⁰	Recipient	Response Time	Appeal Body
Punjab	No fee for 20 pages of the information 2 PKR per extra page charged for 1st 20 pages of the information, Rs. 5 per extra page and Rs. 8 per page if both sides of the page are used	None	Designated Public Information Officer. In case none is designated, head of public body.	14 days	Punjab RTI Commission
KP	No fee to be charged for 1st 20 pages of the information, Rs. 5 per extra page and Rs. 8 per page if both sides of the page are used	CNIC required	Designated Public Information Officer. In case none is designated, head of public body.	10 days	KP
Sindh	There is no application fee. Schedule for cost of photocopying & other material yet to be announced	Copy of CNIC	Public Information Officer. In case none is designated, head of public body.	15 days	Sindh RTI Commission
Baluchistan	Rs. 50 in National Bank of Pakistan which covers first 10 pages of information. Rs. 5 are charged for each extra page.	Copy of CNIC Bank Challan	Head of public body	21 days	Ombudsman
Federal	Rs. 50 in National Bank of Pakistan which covers first 10 pages of information. Rs. 5 are charged for each extra page	Copy of CNIC Bank Challan	Head of public body	21 days	Federal Ombudsman

9. Details of cost schedules of KP and Punjab Acts can be found here <http://www.crti.org.pk/wp-content/uploads/2015/02/CPDI-Comparison-of-Khyber-Pakhtunkhwa-and-Punjab-Schedule-of-Cost-for-providing-Information.pdf>
10. Even though the laws in KP and Punjab do not make the submission of a CNIC copy mandatory, applicants at times find it easier to attach a copy.

Filing RTI Requests

As the previous sections demonstrate, the procedures of filing an RTI application differ slightly from province to province. In Balochistan and for requests dealing with Federal departments, you will first have to submit a fee. The challan form for fee submission can be downloaded from Centre of Peace and Development Initiatives, CPDI website, here: <http://www.cpdipakistan.org/wp-content/uploads/2014/06/challan.pdf>, it is also included as Annex 1. The exact format of the information request will differ, depending upon the jurisdiction where you are filing your information request. The FOI Ordinance 2002 and Balochistan Freedom of Information Act 2005, both require the applicant to share identity details and documents. Thus, applications to departments and other public entities in these regions need to be accompanied by key identifying details and a copy of the applicants Computerized National Identity Card, CNIC. In addition, you will also need to attach proof of payment i.e. a copy of the paid bank challan.

Sample Format

You can use the following format as a sample to develop your own RTI requests.

Information Request Sample Format	
APPLICANT DETAILS	
Name	
CNIC	
Address	
Contact Number	
REQUEST DETAILS	
Public Body Name	
Subject	
Questions	1. 2. 3. 4.
Reasons for information request	
PAYMENT DETAILS	
[Only For FOI Ordinance 2002 and Balochistan FOI ACT 2005]	
Bank Branch	
Receipt No.	
Signature	Date
REQUISITE DOCUMENTS TO ATTACH	
1. CNIC Copy [In Balochistan, KP, Federal Capital & Sindh] 2. Copy of Bank Challan [In Federal & Balochistan]	

The Punjab Transparency and RTI Act 2013 does not make it mandatory for applicants to share CNIC details, so if you want, requests to department in Punjab can be filed without inclusion of CNIC number and copy.

What to do when information requests are rejected or ignored?

First and foremost, it is important to know your rights as defined in the law under which you have filed the information request. By law, the federal government departments and public bodies in Balochistan are required to respond to you in 21 days. In Punjab, the time of response is 14 days, in Sindh it is 15 days and in KP the time of response is 10 days. The law in Punjab also says that if the department is unable to provide the information in the initial 14 days, they can write to inform the applicant that an additional 14 days are required. But, in any case they have to respond, even if to reject the application or inform the applicant about the need for extra time.

Once the given time period is up and you haven't heard back from the public body, or they have rejected your request for information, you can;

- * File an appeal to the federal ombudsman if you have applied under FOI Ordinance 2002 or to the ombudsman if you have applied under Balochistan FOI Act 2005
- * File an appeal to RTI Commission Sindh, KP or Punjab

Framing Effective Requests

Remember that the laws and the culture of promoting freedom of information as a right is fairly new. There is a culture of secrecy that prevails within most government departments. These tips help to make sure that your requests are answered and you receive requisite information in time

- i. Address the request to the relevant PIO or head of department.
- ii. Include any required identity documents.
- iii. Pose a question that is related to that department.
- iv. Pose the question in a clear and concise manner.
- v. Include the dates during which the information or the documents you are requesting were produced.
- vi. Deconstruct complex questions, requests into a set of simpler, easy to understand and difficult to dodge questions.
- vii. Do not attempt to include a vast number of issues and problems within a single request.
- viii. Keep the tone neutral and pose the question to gather information without passing judgment.

RTI for journalists

Now that you know how the RTI law works, it is time to understand what this law means for journalists. The RTI law allows journalists to potentially access data that was previously out of the public domain.

Through these laws, you can now start collecting both data and information that would help you find scoops, do investigative journalism and have a better contextual understanding of issues. For example, previously, journalists reporting on the progress of a project under which a few schools had to be built, would simply have the reports from the ground. There might be speculations about the misappropriation of funds, but without an insider source you will not be able to access financial expenditure details. Under the RTI laws, now you can actually request information about the expenditure details of that particular project.

When filing an RTI request, make sure that you keep a copy of your request and receipt of the registered post that you used to send that request. Both of these will help you if you need to challenge a public's body's response to your request.

THINGS TO REMEMBER

- 1. It is important to identify the relevant public body correctly. Otherwise you simply won't be able to access the information you are after.
- 2. It is good practice to send the information request by registered post, so that you have proof that you sent the request on a specific date
- 3. Always keep a copy of the information request you sent. In case you need to go in appeal, both the registry receipt and the copy of your original request would be needed.



RTI IN ACTION

THE NEWS PUBLISHES A STORY BASED ON
INFORMATION ACQUIRED THROUGH RTI LAW

310 out of 732 Punjab BHUs working sans doctors

ISLAMABAD: Despite claiming success for the construction of mega projects in big cities, the Punjab government has failed to ensure the provision of basic health facilities to almost half of its population, official documents have confirmed.

According to the data collected by this correspondent from 10 districts of the Punjab under the Punjab Right to Information and Transparency Act 2013, out of 732 basic health units in these districts, 310 are operating without any medical officer. The certified information has been provided by the executive district officers (EDOs) Health of these districts under the new RTI law. The information requests were sent separately to all districts of the province on October 21, 2014, but a majority of the districts have failed to respond under the legal time framework. According to the figures, provided by the EDO (Health) Okara, about 62 Basic health units are functioning without a doctor in his district alone. Okara is a key district of the Punjab with a population of 2.2 million people according to 1998 Census. The district had 96 basic health units, but only 34 are blessed with a doctor.

Similarly, in District Sargodha, 78 BHUs have no trained doctor available. The district has a population of around three million and there are 126 BHUs, but only 48 are filled.

According to local residents, people have to walk for miles to reach their respective basic health units only to find out that there is no doctor available in their "official hospital".

Health experts believe many lives could be saved by only deputing qualified medical officer in rural areas. "These medical officers can provide lifesaving support to emergency patients in far flung areas. But the absence of trained doctors in rural areas sometime proves fatal for the poor patients," said senior former director Accidents and Emergency at Pakistan Institute of Medical Sciences (PIMS) Dr Waseem Khawaja.

He said in the absence of doctors, people of rural areas turned to quacks who complicate their medical conditions with dangerous treatments.

"Sometime we receive patients with irreparable damage to their organs because they had been treated by quack in remote areas of the country," Khawaja added. He said one of the reasons for the shortage of medical officers in rural areas is the open merit policy in medical colleges which results in more female doctors.

"In our country, many female doctors refuse to work in rural areas and remote owing to cultural and security reasons. Therefore, there is a need for 50% quota for male students to ensure every village has a doctor available," he added.

According to official data collected from each district separately, Sialkot district has 88 sanctioned posts for doctors in BHUs, but 33 units have no doctors available as their posts are vacant.

In Muzaffargarh, out of 71 sanctioned posts of medical officers in BHUs 12 are vacant, in Kasur 25 BHUs are operating without doctors.

In Bhakkar, 19 BHUs have no doctor while the district has total 39 health units. Mandi Bahauddin has 49 BHUs, but 18 have no doctor available.

In Toba Tek Singh, 8 out of 70 BHUs have no doctors. In Sahiwal, vacant posts of doctors in BHUs are 30. In Chiniot, only 11 BHUs have doctor while 25 are almost non-functioning owing to absence of doctors.

Another story published in The News, based on information collected through RTI requests revealed that “According to certified data collected from 23 districts of the two provinces [KP and Punjab] through their respective Right to Information (RTI) laws, a large number of oxygen cylinders, ultrasound, ECG and CT scan and X-Ray machines are non-functional in their district headquarters (DHQ) hospitals, causing severe problems for the poor residents of these areas who are unable to afford private medical care. The certified information also made shocking revelation that district headquarters (DHQ) hospital of Pakpattan, which caters to over 1.2 million people, does not have even a single functional X-ray machine.”

As these stories demonstrate, RTI opens up a whole new treasure chest of information that can lead reporters towards exclusive stories, particularly stories dealing with corruption, financial misappropriation and governance related issues.

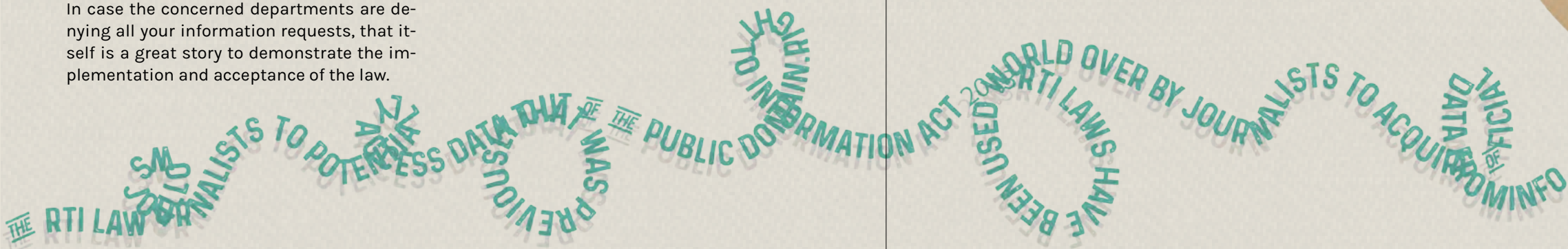
In case the concerned departments are denying all your information requests, that itself is a great story to demonstrate the implementation and acceptance of the law.



Exercise

Go back to the story you looked at in the exercise about data sources (page 27). Monsoons are about to begin and you will be covering the spread of dengue fever in your area. Think about the information that can help you better prepare for the coverage;

- i. Which departments can you send an RTI request to?
- ii. Which hospitals can you send RTI requests to?
- iii. What kind of data will you ask for?



CHAPTER SIX

THE QUANTIFICATION OF INFORMATION

REPORTERS are usually in possession of a tremendous amount of information. This information comes to them regularly through their official and other sources. At times this information includes some data, like budgetary data, data about stocks and markets and numbers of affected people during natural disasters or terrorism incidents. But, most of the times, this information is in descriptive formats and not countable. In this chapter, you will learn how to look at information from the quantification lens i.e. how to identify countable elements within information sets and how to set up databases to manage and analyze this data.

In this manual, we are not concerning ourselves with information theories that deal with quantification on an academic level. Since this manual is designed as an easy to use resource from which aspiring data journalists can easily benefit, we will focus only on how quantification will apply within news rooms. For that it is enough to understand that in social sciences, we sometimes seek to quantify, or to bring in numerical countable form, values that are originally qualitative. For those of you who have done surveys, you might remember putting qualitative variables on a scale (like asking people if they agree, disagree, strongly agree etc. with a statement). We can see similarly qualitative data quantified through a set of indicators like the world happiness index or the human development index that create quantitative, countable indicators to measure variables that are essentially qualitative or not directly countable.

This is what we aim to do with some information you, as reporters gather.

Data quantification and management

Think of quantification as formal management of a story you file again and again. For example, crime reporters among you essentially file the same story every day – crimes that happened in the city – the story differs only in the nature of the crime, the names of the victims change, the circumstances in which the crime was committed change, the status of investigation changes, but in essence, what they are reporting is how, where and who violated the rules set by Pakistan Code of Criminal Procedure, PCP, who became the victim of this violation and who / how is

trying to find the culprit. This means that the crime reporter is reporting on the 5Ws and H of all violations of the PCP.

So, they are basically looking at the same set of variables every day.

To quantify this or any other information, that is what you have to look at – which are the stories that are repeating themselves within your beat with a simple change in the value of the variable.

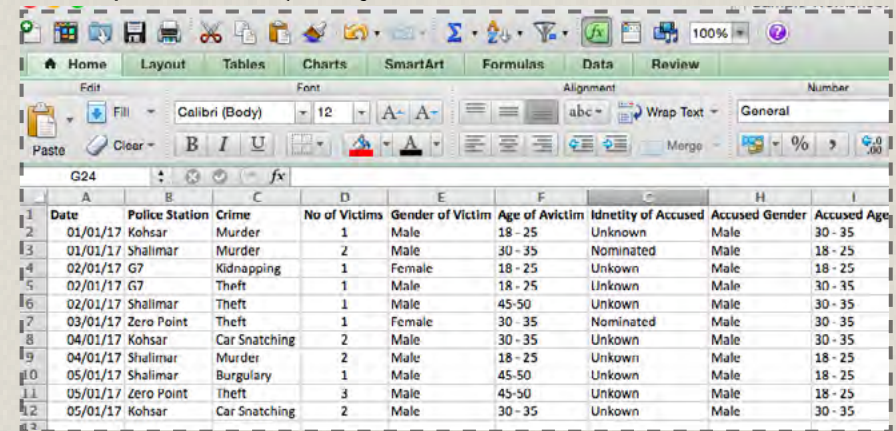
Going back to the example of the crime reporter, here is a basic list of variables that present themselves in each story

1. Date of the crime
2. Nature of crime
3. Area of the crime
4. Jurisdiction
5. Demographical details of the victim
6. Demographical details of the accused
7. Impact of the crime [like loss of life, property, injuries etc.]
8. Response from law enforcement
9. Status of investigation

What data quantification entails is to store this data in a manner that you can update the specifics about each crime that is reported to you in a quantifiable database. So, if you simply make an excel sheet and start entering basic details of all crimes reported from your city every day, you would soon have an extensive database that you can use to create investigative, data based stories.

The advantage of creating your own databases is that you are simply using the information you gather in your daily routine and do not even have to look for an additional data source. Keeping your information in the form of a database means that you will be able to use and reuse the same information in a more effective manner. Once the database is created, you can map it, track it, create infographics, multimedia and essentially play with it in various different ways.

A sample dataset to quantify crime related data



	A	B	C	D	E	F	G	H	I
1	Date	Police Station	Crime	No of Victims	Gender of Victim	Age of Victim	Identity of Accused	Accused Gender	Accused Age
2	01/01/17	Kohsar	Murder	1	Male	18 - 25	Unknown	Male	30 - 35
3	01/01/17	Shallimar	Murder	2	Male	30 - 35	Nominated	Male	18 - 25
4	02/01/17	G7	Kidnapping	1	Female	18 - 25	Unknown	Male	18 - 25
5	02/01/17	G7	Theft	1	Male	18 - 25	Unknown	Male	30 - 35
6	02/01/17	Shallimar	Theft	1	Male	45-50	Unknown	Male	30 - 35
7	03/01/17	Zero Point	Theft	1	Female	30 - 35	Nominated	Male	30 - 35
8	04/01/17	Kohsar	Car Snatching	2	Male	30 - 35	Unknown	Male	30 - 35
9	04/01/17	Shallimar	Murder	2	Male	18 - 25	Unknown	Male	18 - 25
10	05/01/17	Shallimar	Burglary	1	Male	45-50	Unknown	Male	18 - 25
11	05/01/17	Zero Point	Theft	3	Male	45-50	Unknown	Male	18 - 25
12	05/01/17	Kohsar	Car Snatching	2	Male	30 - 35	Unknown	Male	30 - 35

This is just a sample dataset with dummy data that a crime reporter can create to quantify and manage the information that he or she is gathering on a daily basis. Other elements can be added to this data set as well. For example, First Information Report, FIR number, the time of the crime, whether there were any witnesses, whether statements have been recorded etc. So, quantification of data allows you to create a dataset using the information you have, that can lead you to new stories, help you identify trends, support the creation of interesting infographics and new formats of storytelling, all without having to do the additional exercise of finding a new data source. The database you make from your own data will also allow you a certain degree of exclusivity.

Some basic rules that you have to follow;

i. Think carefully about the inclusion and exclusion of categories or variables in your data set. There is no right or wrong formula. You just have to think about the final story you are trying to get out of your data. There is no need to add additional categories if they will add nothing to your story. Similarly, keeping the focus too narrow will take away the depth of your story. For example, another category that you can include in the crime data might be the weather/temperature on the day of crime. But do you really think it will add anything to the analysis? Yet another thing that can be added to the same data set might be the law under which the FIR has been registered. Which categories you add de-

pends on your own preferences, but the rule is to think about all the possibilities at the time you are creating the database.

ii. Perhaps the most important rule is to make sure that you limit the possible responses in as many columns as possible. Having an unlimited set of possible responses will make the dataset unsuitable for mathematical analysis. There would be some categories/variables or columns in which every entry would be unique i.e. the possible responses unlimited. But in most cases, you will be able to have a restricted set of responses. For example, in the crime dataset we just saw as a sample, a column might be added to document the name of the victim. Now, in this column you cannot possibility limit which names can be added in response. But almost all the other columns would have limited set of responses. For example, if this crime data is from Islamabad, the names of police stations is a limited pool. Similarly, there are a limited number of crimes defined in PCP and other laws. You can limit the responses in age column by making sets of defined ranges [10 - 15, 15 - 20, 20 - 25 etc.]. Limiting the possible responses makes the dataset viable for quantitative analysis. If all the entries in all the columns were unique, there would be no easy way to map any trends or find any insight into what the data means. You will understand this better when you are doing exercises with spreadsheets.

- iii. Within the limited set of responses, you define for each category, add a specific response to indicate the nil value. For example, in most registered FIRs the case is registered against unknown suspects. Thus, you need to have a specific response 'unknown' in the accused details columns. If you don't add it, you might have to leave a lot of responses blank which will affect the final analysis.
- iv. It is good practice to add an optional column titled additional comments, in which you can briefly indicate any interesting or unique aspects of an individual entry.

A Sample Dataset

Here we will introduce you to a sample dataset, that has been created using the quantification technique, and that you will be using for technical exercises as you go further in this manual. This is the data representing casualties of media workers. Before we move further, you need to download the data set from here.

Name	Design	Org	Year	Province	City	Type	Targeted	Medium	Threatened
Chahid Mujahid	Columnist	Akhbar-e-Jehan magazine	2008	Balochistan	Quetta	Shot dead	Yes	Print	Yes
Khadim Hussain Sheikh	Reporter	Daily Khabran	2008	Balochistan	Lasbela	Shot dead	Yes	Print	Uncertain
Talash Vohra	Reporter	Karwan daily	2008	Balochistan	Lasbela	Shot dead	Yes	Print	Uncertain
Wasi Ahmad Qureshi	Reporter	Daily Balochistan Express	2009	Balochistan	Khudat	Shot dead	Yes	Print	Uncertain
Malik Araf	Camerman	Samaa TV	2010	Balochistan	Quetta	Suicide blast	No	Broadcast	No
Ejaz Rasani	Reporter	Samaa TV	2010	Balochistan	Quetta	Suicide blast	No	Broadcast	No
Abdul Hammed Hayat	Reporter	Daily Inshah	2010	Balochistan	Gwadar	Shot dead	Yes	Print	Yes
Faz Khan Saeed	Reporter	Sawa News TV, Daily Zamana	2010	Balochistan	Khudat	Shot dead	Yes	Print - Broadcast	Uncertain
Iqbal Nisar	Sub Editor	Baloch magazine	2011	Balochistan	Turnat	Shot dead	Yes	Print	Uncertain
Abdool Rind	Reporter	Daily Eagle	2011	Balochistan	Turhat	Shot dead	Yes	Print	Uncertain
Muneeb Shakir	Reporter	Online news agency	2011	Balochistan	Khudat	Shot dead	Yes	News Agency	Uncertain
Javed Nasser Rind	Sub Editor	Daily Tower	2011	Balochistan	Hala	Shot dead	Yes	Print	Uncertain
Abdul Rasool Gul	Reporter	Express News TV	2012	Balochistan	Turhat	Shot dead	Yes	Broadcast	Uncertain
Abdul Qadir Majidi	Reporter	Wahab TV	2012	Balochistan	Wahab	Shot dead	Yes	Broadcast	Uncertain
Abdul Haq Bano	Reporter	ARY News TV, Daily Awan, Radio Pakistan	2012	Balochistan	Khudat	Shot dead	Yes	Print - Broadcast	Uncertain
Behmdad Khan	Reporter	Daily Quana, AFP, Daily Inshah	2012	Balochistan	Pingir	Shot dead	Yes	News Agency - Newspaper	Uncertain
Imran Shukh	Camerman	Samaa TV	2013	Balochistan	Quetta	Bomb blast	No	Broadcast	No
Saif Ur Rehman	Reporter	Samaa TV	2013	Balochistan	Quetta	Bomb blast	No	Broadcast	No
Muhammad Usaf	Camerman	NKI news agency	2013	Balochistan	Quetta	Bomb blast	No	News Agency	No
Muhammad Ahmed Afridi	Reporter	Daily Inshah	2013	Balochistan	Qalat	Shot dead	Yes	Print	Uncertain
Imdad Memon	Reporter	Online News Agency and ARY News TV	2014	Balochistan	Quetta	Shot dead	Yes	News Agency - TV	Uncertain
Abdul Rasool	Reporter	Online News agency	2014	Balochistan	Quetta	Shot dead	Yes	News Agency	Uncertain
Zafarullah Jaleel	Reporter	Daily Inshah	2015	Balochistan	Jaffarabad	Shot dead	Yes	Print	Uncertain
Sajid Taseel	Reporter	Daily Journal	2006	Federally Admin Minerva	Quetta	Shot dead	Yes	Print	Uncertain
Ali Shah	Reporter	The Nation daily	2006	Federally Admin Wana	Shot dead	Uncertain	Print	Uncertain	Uncertain
Mir Nawaz	Reporter	AP news agency	2007	Federally Admin Quetta	Shot dead	Yes	News Agency	Uncertain	Uncertain
Nasir Afridi	Reporter & President Tribal Union of Journalists	Daily Khabran	2006	Federally Admin Dera Adam	Killed in Crop	No	Print	No	No
Hayatullah Khan	Reporter	ATP news agency	2006	Federally Admin North Wazir	Shot dead	No	News Agency	Yes	Yes
Naveen Hashim	Correspondent, Vice President Tribal Union of Journalists	Daily Pakistan	2007	Federally Admin Bajaur	Roadside bla	No	Print	No	No
Muhammad Ibrahim Khan	Reporter	Express News	2006	Federally Admin Bajaur	Shot dead	Yes	Broadcast	No	No
Hassanullah Khan	Reporter	Shamshad TV (Afghanistan)	2009	Federally Admin Jandul	Shot dead	Yes	Broadcast	Yes	Yes
Azmat Ali Bano	Reporter, Camerman	Samaa TV	2010	Federally Admin Chitral	Agri-Suicide blast	No	Broadcast	Yes	Broadcast
Rameez Khan	Reporter	Wazir TV	2010	Federally Admin Mohmand	Agri-Suicide blast	No	Print - Broadcast	No	No
Abdul Wahab	Reporter	Express News	2010	Federally Admin Mohmand	Agri-Suicide blast	No	Broadcast	No	No

Journalist Killings in Pakistan Dataset

This is a screenshot of the dataset that you will download, you will need Microsoft Excel to open and use it.

This dataset quantifies a tragic story that you all are familiar with - the killings of journalists in Pakistan. Even if you do not report crime, your association with the journalist community in Pakistan automatically makes you a part of this story. Journalist killings in Pakistan have continued for the past 17 years. Every year, we have a new set of tragic killings to report. This data set keeps track of a few variables that repeatedly occur in the reporting of this story, namely;

- 1. The name of the journalist
- 2. The designation
- 3. Organization
- 4. The medium he worked for
- 5. Year of killing
- 6. City
- 7. Province
- 8. Nature of killing
- 9. Whether he was targeted
- 10. Whether he was being threatened or not

Individually, all these killings have unique details. But, put together in a database like this, they allow you to see the larger picture and map the trends that you might miss when looking at the reporting of individual targeting.

THINGS TO REMEMBER

1. The only way you can actually do analysis on a dataset is when you have a limited set of responses against each category or variable

2. Have a specific response for entering a nil value – for example unknown, not applicable, does not apply etc. This will allow you to document nil values rather than leaving blank cells in your database. Blank cells complicate data analysis.

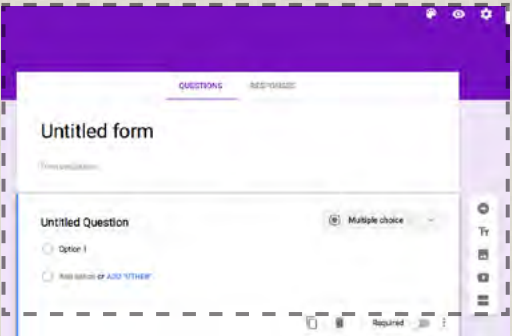
3. Try making sure that each column includes just one variable, for example break down the month and year in two columns. This will help in data analysis later.

A tool for easy data management

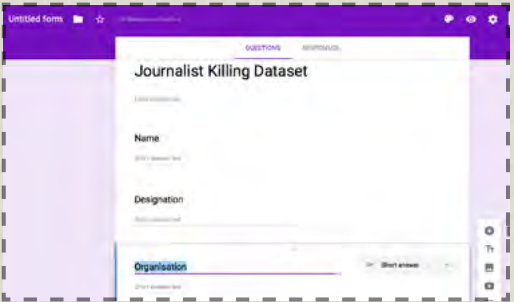
If you are like us, looking at spreadsheets makes you a little nervous. We have grown up associating spreadsheets and excel with finance and accounting and thus, as journalists we might be a apprehensive about using them. Now, when it comes to data entry and management, particular when you are maintaining databases of your own, you can use a tool that makes it easy for you to update the data without actually having to open the spreadsheet. (Regardless, you will be working plenty with spreadsheets in any case, as you will see in Chapter Eight.)

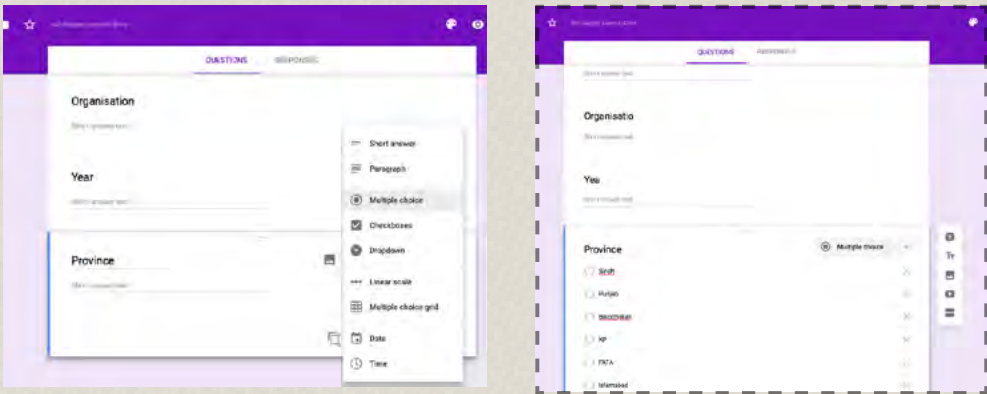
The tool we are talking about is Google Forms. This is free software that is accessible through Google Drive. It allows you to create simple forms, like surveys, which you can use to enter data into your spreadsheet. On the backend, Google manages the spreadsheet and generates basic analysis on the data for you as well. The best part? Using Google Forms is as easy as it gets. Here is how you create forms for your datasets;

- i. Sign into your Google account and go to drive.google.com
- ii. Click ‘New’ at the top left. You will see a list of apps, if Google Forms is showing in the list, click on Google Forms. If it is not showing, click ‘Connect more apps’, search for Google Forms, click ‘Connect’.
- iii. Once you have connected Google Forms, you will get this scree

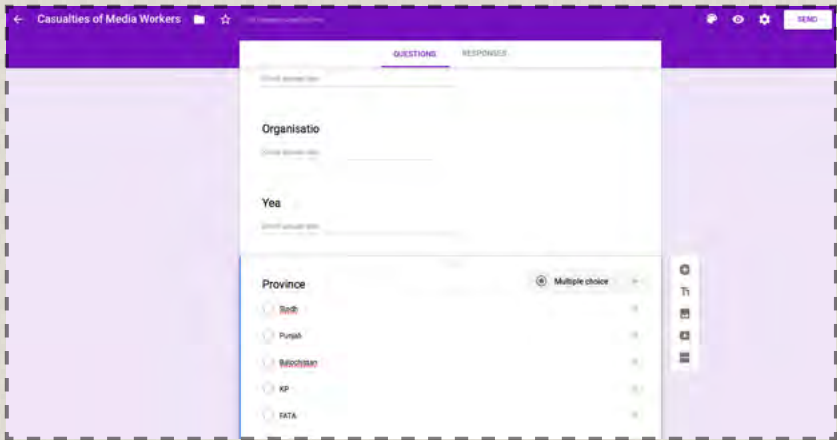


- iv. Add the name of the database and other details in the description spaces. To demonstrate, we will use the details from the Journalist Killings Dataset you saw earlier in this chapter.
- v. Use the plus symbol on right to add more categories. If you are expecting others to fill in the data, you can use full questions to help them understand what exactly is the response you are looking for. If you will be updating and maintaining the dataset yourself, it is enough to simply add the title of the categories.



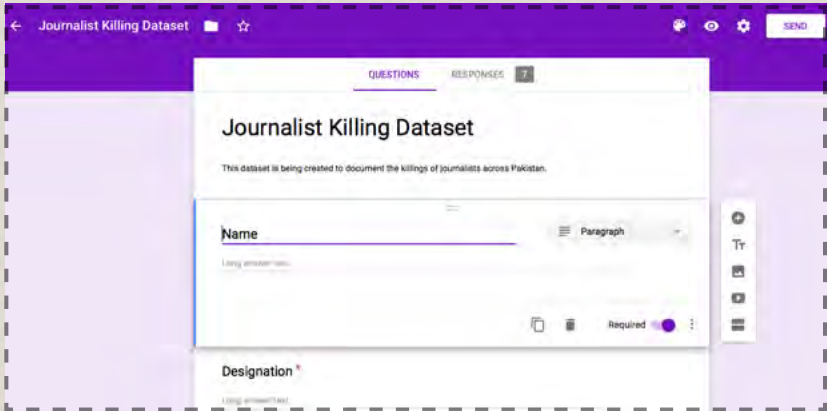


On clicking the arrow, you will see a list of options, which shows you the kind of responses you can get. In this case, we will select the ‘Multiple Choices’ option to list out the provinces.



Once your dataset is ready scroll back up and click send. You can also preview how it looks by clicking the eye icon near send and edit to correct any mistakes you find in preview. Clicking ‘Send’ will give you the options to send by email, get a link to share or get an embeddable form.

To make sure that none of the questions are left unanswered by mistake, click ‘Required’ at the bottom of each question.

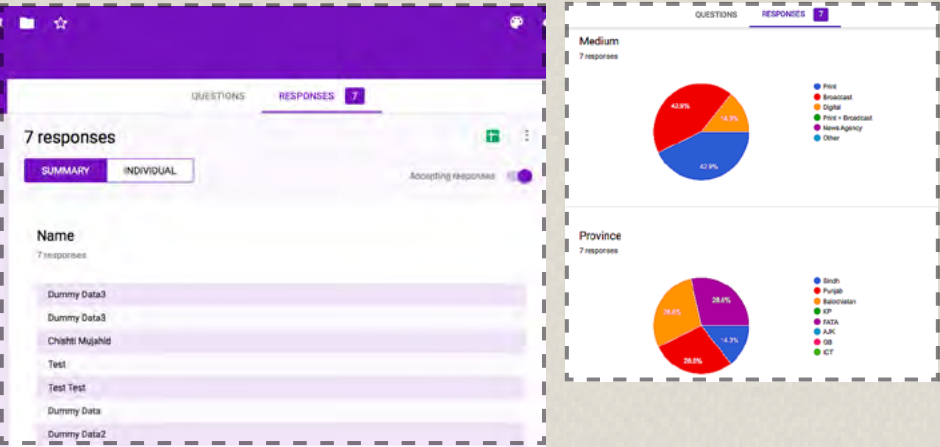


The red star that appears in front of Designation in this screenshot shows that it is a required question. Now, if you are filling the form (or anyone else is filling it for you), you will not be able to submit it if a question is accidentally left unanswered. The blank question will be highlighted for you to respond to.

Here is the link to the form we just created - <https://goo.gl/forms/s5X80VibISGstv952>

You can put this URL in your browser, or just click the link if you are using the digital version of this manual to see the form. You can add dummy data in the form as well.

vii. And finally, we come to the dataset this form will help you create. At the top of the form, in the original (edit) window, click the ‘Responses’ tab.



Here is what you will see in the response tab – a summary of all responses, including some basic graphs that analyze the data for you. But, most importantly, you will see a small green icon on the top right. This is the icon for spreadsheets. When you click this icon, Google will ask you if you would like to create a spreadsheet linked to your form, click ‘Yes’, and you have an automatically generated spreadsheet that will be updated every time you enter new data through the form. You can view the spreadsheet that is linked to the test form we just created for you here - <http://bit.ly/2nJS4Ek>.

Go back to the form, enter dummy data and refresh the spreadsheet to see your dummy data appear.

This spreadsheet can be linked directly to Google’s visualization tools, including mapping tools that we will learn further ahead or it can be downloaded for analysis in Excel, which you will learn in Chapter Eight.

SECTION
THREE

**MAKING
SENSE
OF DATA**



CHAPTER SEVEN

INTERVIEWING THE DATA SET

THE data is now in your hands. Paul Bradshaw, a famed data journalist and trainer, says that “data can be the source of data journalism, or it can be the tool with which the story is told—or it can be both. Like any source, it should be treated with skepticism; and like any tool, we should be conscious of how it can shape and restrict the stories that are created with it”. In this chapter, we discuss the dissection of data sets for journalistic purposes.

Know the data

Knowing the data before you proceed with the analysis is extremely important. You need to be sure of its source, its authenticity and its purpose. You need to understand what it represents and what it does not. And finally, you need to see if there is a critical component missing and how the existing data fits with other available data sets.

Confused? Let us make it simple.

Let's start by interviewing our existing dataset. We can ask the following potential questions:

1. **Who made it?** - This simply refers to the source of the data. The question helps us identify whether this data was/is created and hosted by a government body, an NGO or any other public or private institution, which helps us assess its authenticity. For instance, data acquired for the official website of Federal Bureau of Revenue, FBR could be referred to as ‘public data, hosted by a credible government body’ and hence

authentic. Whereas data acquired from a ‘private’ source, with no means of verification, might bring in questions of its credibility.

2. **Why did they make it?** - This refers to the purpose the data aims to serve. The question helps us identify what you can and/or cannot do with the data. For instance, the taxation data available on FBR's website was made to ‘keep a record of tax payers’. Using the FBR dataset you can ‘find out if a citizen is a tax payer’ but you cannot see ‘if there exists a criminal record for that person’.
3. **When did they make it?** - This perhaps is the most important question. Goes without saying the data is only relevant if it serves the timeline of your journalistic project.
4. **How did they make it?** - Knowing the methodology of data collection/processing is extremely important and can impact figures. For instance, in the case of

our original dataset (casualties of media workers), 111 media workers are listed as ‘individuals who have lost their lives in the line of duty’. Another organization maintaining a similar database might only record figures of ‘media workers killed for their journalistic work’ (excluding the deaths in the field incidents, such as bomb blasts etc.) in which case the total figure would only be 74.

5. **What is missing?** - This question helps us find gaps in the data. Identifying these dodgy elements helps us improve our journalistic product. For instance, let's assume the ‘year’ column is missing in our original dataset (casualties of media workers). The resulting analysis would completely change the angle of your data story.

And finally,

6. **What other data sets could you use, in supplement to your existing data?** - This requires a little out-of-the-box thinking with context to your original journalistic product. For instance, in the case of our original dataset (casualties of media workers), you could try and analyze it with ‘elections timeline’ which can help you identify if the killings peak before, during or after general or by-elections. Similarly, you can analyze it with the timeline of ‘governance terms’, which will help you identify the number of casualties in each political era. Sky is the limit once you get the hang of it!

THINGS TO REMEMBER

1. Always be sure to use data from credible, ideally public sources. Always be sure to credit the source.
2. When you come across a good online data resource, be sure to ‘bookmark’ it in your browser.
3. Always read the ‘data collection methodologies’ mostly available with online datasets.
4. If you find a critical element missing from a dataset, try getting in touch with the author(s). It could be a technical glitch.
5. Try to download and save the datasets (Excel and/or PDF files) on your personal computer as much as you can, in contrast to using online resources only.



Exercise

Here are links to a few datasets that are available online;

- i. Islamabad Police Crime Statistics - <http://islamabadpolice.gov.pk/CrimeStatistics>
- ii. Punjab Police Honor Killing Statistics - <https://www.punjabpolice.gov.pk/honour-killings>
- iii. Sahil Report Cruel Numbers - <http://sahil.org/wp-content/uploads/2016/04/Final-Cruel-Numbers-2015-16-03-16.pdf>

Interview the data available on these sources. Do you find any story ideas that are worth working on? Can these figures be used in ways that are different than the coverage you will usually see on TV or in papers?

Deep Diving into Data

1. So now that we have answered basic questions about our dataset, we are now ready to move on to a more technical exercise – data analysis. What is data analysis? Data analysis is the process through which you examine various components of any given dataset, with an aim to uncover the story within. Let us take an example using our original dataset.

This is how the data in its raw form looks like:

Name	Design	Year	Reported Date	Province	City	Type	Targeted	Killed	Medium
1. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
2. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
3. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
4. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
5. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
6. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
7. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
8. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
9. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
10. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
11. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
12. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
13. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
14. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
15. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
16. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
17. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
18. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
19. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
20. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
21. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
22. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
23. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
24. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
25. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
26. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
27. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
28. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
29. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
30. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
31. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
32. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
33. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
34. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
35. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
36. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
37. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
38. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
39. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
40. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
41. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
42. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
43. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
44. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
45. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
46. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
47. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
48. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
49. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
50. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
51. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
52. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
53. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
54. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
55. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
56. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
57. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
58. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
59. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
60. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
61. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
62. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
63. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
64. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
65. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
66. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
67. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
68. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
69. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
70. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
71. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
72. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
73. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
74. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
75. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
76. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
77. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
78. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
79. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
80. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
81. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
82. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
83. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
84. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
85. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
86. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
87. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
88. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
89. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
90. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
91. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
92. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
93. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
94. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
95. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
96. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
97. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
98. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
99. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper
100. Faisal Qureshi	Web Editor	2011	07-Oct-11	Punjab	Lahore	Tortured to death	Yes	4	Newspaper

Fig 7.1 – Casualties of Media Workers

A little too complicated, no? Not really! You will learn to dissect and deconstruct data in the coming chapters. It's easier than you think!

Before we proceed to more technical aspects of data dissection, do you know what is 'visual analysis of data'?

To make more sense of data in Fig 7.1, and after asking the basic questions about the dataset, let us now proceed to know more about this data. We have a few more questions, but mind you, these questions are 'not' about the 'dataset' but about the 'data' itself.

- 1. What journalistic queries can this data answer or what questions does it pose? - This will help us identify the basic story or stories in the data. For instance, in the case of our primary dataset (casualties of media workers), existing data can tell us the total number of journalists who died in the line of duty, most dangerous city for media workers, most dangerous year and so on.
- 2. How can this data be summarized? - Remember, the data will only make sense if you are able to summarize it and/or perform other basic functions on it, such as count, sum, average, percentage etc. For instance, our default dataset is what we call 'a record based dataset'. It doesn't contain any values or total, but just the listing of information. To be able to calculate total number of casualties, we will have to perform a 'count' function. We will learn all of it and more in chapters to come].

Visual analysis of data is a process through which you can deconstruct components of a complex dataset for ease of your understanding, through simple tools and techniques. Through the visual analysis, you get a sense of the story that particular dataset is telling you. It is then you proceed with more technical dissection of data through a spreadsheet software.

- 3. What other data would help me ask more questions? - The data is not only expected to answer questions, but 'pose' questions as well – questions which you'll have to find answers for through research, web searches etc. This will point you to any external data that you may require to make sense of your story. For instance, in the case of our default data, the 'demographics' seem to be missing.
- And lastly,
- 3. What other data would help me understand this data better? - To better understand the data, you often need to analyze it with respect to external data. The trends that you identify from your dataset can potentially point towards a story, if analyzed with another dataset. For instance, in our dataset (casualties of media workers) the spikes in deaths could be related to ongoing military operations, elections and other factors – but for that you need a timeline of external events, which is not present in this dataset.

THINGS TO REMEMBER

For information on external events (such as, in this case, the dates of military operations in Swat and Waziristan, the dates of general elections and related events) could easily be found by a simple Google search. Be sure to use authentic sources only – ideally a reputed news publication. Crowd sourced information such as 'Wikipedia' is NOT a reliable for journalistic purposes.

What's next?

To be able to analyze data and summarize it for journalistic products, you need to be able to perform simple functions such as 'sort' or 'filter' to answer a particular question that you (as journalist) might have. For instance, in our original dataset (Casualties of Media Workers), you might want to know 'How many journalists died in the line of duty in 2007?' or 'What province has suffered most losses?' We will learn these basic functions in the next chapter. But before that, you need to raise questions. Here is an exercise that will help you ask more questions (the answers for which we will try to find in the next chapter.



Exercise

Open the dataset given with this manual and try to think of the five most important questions that will help you produce a data story on this issue. Following are some sample questions:

- 1. What is the total number of journalists who have died in the line of duty (from 2000 to 2015)?
- 2. Which is the deadliest role for a media worker (reporter, cameraman etc.)?
- 3. Which is the deadliest province for media workers (highest number of casualties in a province)?
- 4. And, how many of those who died in the line of duty were targeted?

CHAPTER EIGHT

SPREAD SHEETS, THE DATA JOURNALISTS' SECRET WEAPONS

SPREADSHEETS a word that journalists traditionally associate with accounting, finance and statistics; in other word a tool that most of us have learned to steer away from. For data journalists however, spreadsheets are a effective tool of trade. Spreadsheets are an essential part of the Open Data movement and across the world spreadsheet format is favorably seen and utilized. In this chapter, we will learn to use spreadsheets, particularly Microsoft Excel to conduct basic analysis on datasets. This is a hands-on, technical session and so before we proceed before make sure you have following:

Prerequisites for this Chapter

- 1. Basic skills to use computer applications and software.
- 2. A Windows or an Apple computer with Microsoft Excel installed.
- 3. Sample data file, titled 'Casualties of Media Workers' (also available on a digital media with this manual).
- 4. And most importantly, eagerness to learn new technologies!

What are spreadsheets?

Spreadsheets are interactive computer applications often used to store and analyze data in a tabular format (in the form of tables).

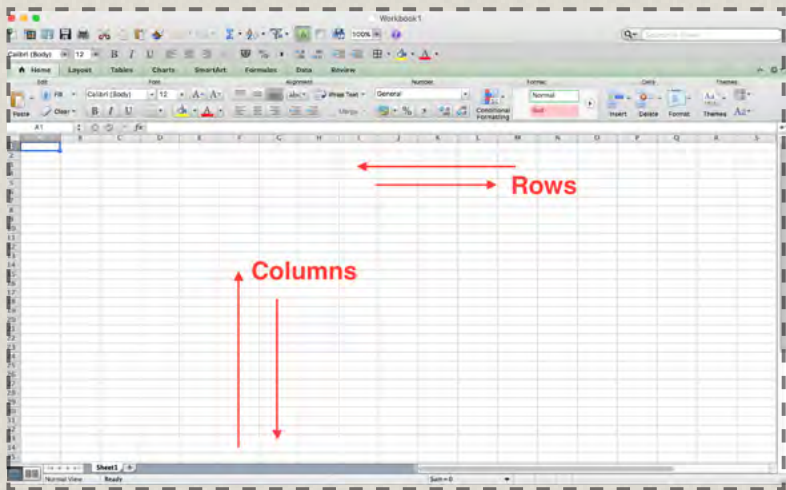


Fig 8.1 – Interface of a traditional spreadsheet software

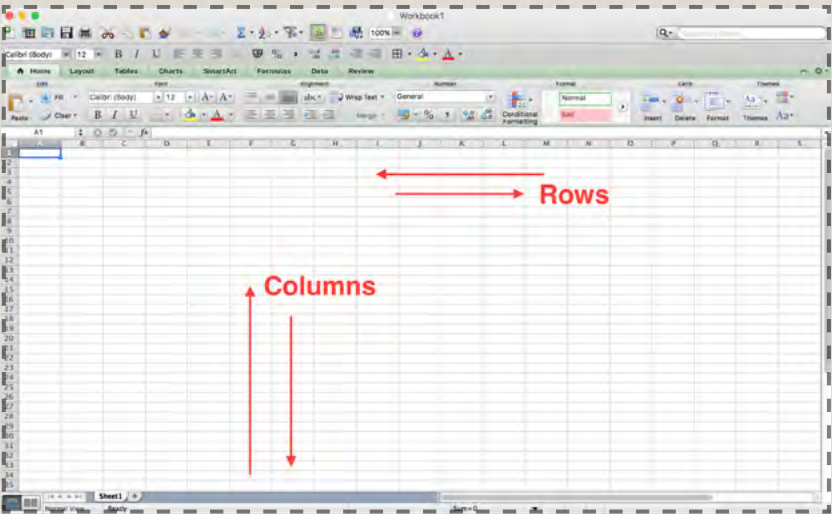


Fig 8.2 – Cells

Tables, as you know have rows and columns (as shown in figure 8.1). Each row and column can be selected individually for various purposes including but not limited to, sorting the file in an ascending or descending order, performing calculations and filtering functions. Rectangles on the cross-section of rows and columns are referred to as ‘cells’ (as shown in figure 8.2). Each cell can essentially be treated as a separate entity and has a unique identification number. A cell can store data in multiple formats including but not limited to text, number, date and currency. **For demonstration purposes, we will be using Microsoft Excel (as a spreadsheet software) in this manual.**

But do not worry, you are not getting a lesson in accounting today!

Let us see what we can do with Microsoft Excel for data journalism. There are two stages of data analysis for journalism; the first one is visual analysis (defined in Chapter Seven) which we will practice in a just a bit. The other stage is slightly more advanced and requires a deeper understanding of Microsoft Excel, covered in details in the next chapter.

Using Microsoft Excel for visual data analysis

We have a dataset, now what? The first step of understanding any dataset, complex or simple, is a basic visual analysis, which helps us understand various components of a dataset and lets us answer basic journalistic questions about that particular data. It also tells us what other information and/or data is required to produce a good data story based on the particular dataset that we have acquired. The good news is, that a basic visual analysis is extremely easy and does not require an advanced knowledge of Microsoft Excel. Remember that visual analysis is only to get a sense of data so you can make better use of it. During the process, you do not need to worry about finding out the ‘exact’ figures/trends. We will cater to that in the next chapter.

To better demonstrate this, let us start with our original dataset (Causalities of Media Workers). In the exercise in Chapter Seven, we listed some sample question, the answers of which will help us narrate our story. Let us try and find answers to those questions through Microsoft Excel.

Question 1: Total number of journalists who have died in the line of duty?

This is what our data looks like:

Name	Design	Org	Year	Reported Date	Province	City	Type	Targeted	Killed	Medium
Tariq Karim	Assistant Editor	Daily Kanish	2012	09-May-12	Sindh	Karachi	Shot dead	Yes	8	Newspaper
Murtaza Rizvi	Open	Open	2012	08-Apr-12	Sindh	Karachi	Turned to death	Yes	14	Magazine
Zubair Ahmed Mujahid	Bureau Chief	Daily Jang	2007	23-Nov-07	Sindh	Mirpurkhas	Shot dead	Yes	2	Newspaper
Imdad Memon	Bureau Chief	Online News Agency and ARY News TV	2014	28-Aug-14	Balochistan	Quetta	Shot dead	Yes	8	News Agency + TV
Saeed Saleem Shahzad	Bureau Chief	Asia Times Online	2011	29-May-11	Punjab	Mandi Bahaudin	Turned to death	Yes	1	Newspaper
Samuel Pearl	Bureau Chief	Wall Street Journal	2008	21-Feb-08	Sindh	Karachi	Shot dead	Yes	14	Newspaper
LM Shalizi	Chief Editor	Daily London News	2008	29-Nov-08	Punjab	Lahore	Shot dead	Yes	4	Newspaper
Mahmood Syal	Correspondent	Online news agency	2006	14-Sep-06	Kuber Pakhtunkhwa	Dera Ismail Khan	Shot dead	Yes	3	News Agency
Mahmood Syal	Editor	Daily Nisr	2007	23-Jan-07	Sindh	Kharpur	Shot dead	Yes	3	Newspaper
Tariq Aalam	News Editor	Daily Pakistan	2013	16-Apr-13	Kuber Pakhtunkhwa	Peshawar	Suicide blast	No	8	Newspaper
Muhammad Ismail Malik	Resident Editor	PII news agency	2013	30-Oct-08	ICT	Islamabad	Shot dead	Yes	3	News Agency
Saghar Hussain Karam	Sub Editor	Daily Morning	2009	20-Feb-09	Sindh	Karachi	Shot dead	Yes	14	Newspaper
Imdad Memon	Sub Editor	Daily Jang	2011	09-Nov-11	Balochistan	Hub	Shot dead	Yes	1	Newspaper
Imdad Memon	Sub Editor	Baloch magazine	2011	09-Jan-11	Balochistan	Turbat	Shot dead	Yes	3	Magazine
Abdul Basit Beshir	Sub Editor	Daily Jang	2013	21-Aug-13	Sindh	Karachi	Turned to death	Yes	14	Newspaper
Faisal Qurashi	Web Editor	The London Post	2011	07-Oct-11	Punjab	Lahore	Turned to death	Yes	4	Newspaper
Imdad Memon	Sub Editor	SM Digital (UK based)	2006	06-Jul-07	ICT	Islamabad	Killed in Crossfire	No	3	TV
Muhammad Aijaz	Camerman	ARY One World TV	2007	18-Oct-07	Sindh	Karachi	Suicide blast	No	14	TV
Muhammad Aijaz	Camerman	Express News TV	2009	04-Jan-09	Kuber Pakhtunkhwa	Dera Ismail Khan	Suicide blast	No	14	TV
Malik Arif	Camerman	Samaa TV	2010	16-Apr-10	Balochistan	Quetta	Suicide blast	No	8	TV
Imdad Memon	Camerman	Samaa TV	2013	30-Jan-13	Balochistan	Quetta	Suicide blast	No	8	TV
Muhammad Iqbal	Camerman	NBA news agency	2013	30-Jan-13	Balochistan	Quetta	Suicide blast	No	8	News Agency
Mir Nawaz	Sub Editor	AP news agency	2009	07-Feb-09	Federally Administered Tribal Areas	Wana	Shot dead	No	2	News Agency
Muhammad Syal	Camerman	KTN TV	2006	29-Jan-06	Sindh	Larkana	Killed in Crossfire	No	2	TV
Chaudhry Muhammad	Columnist	Akhbar-e-Jehan magazine	2008	09-Feb-08	Balochistan	Quetta	Shot dead	Yes	8	Magazine
Noor Mahmood	Correspondent, Vice President	TDI Daily Pakistan	2007	02-Jun-07	Federally Administered Tribal Areas	Bajaur	Roadside Blast	No	2	Newspaper
Imdad Memon	Camerman	Express News	2014	19-Jan-14	Sindh	Karachi	Shot dead	Yes	14	TV
Muhammad Aijaz	Camerman	Express News	2014	19-Jan-14	Sindh	Karachi	Shot dead	Yes	14	TV
Muhammad Aijaz	Camerman	Express News	2014	19-Jan-14	Sindh	Karachi	Shot dead	Yes	14	TV
Muhammad Aijaz	Camerman	ARY News	2014	21-Sep-14	Sindh	Karachi	Shot dead	Yes	14	TV
Arshad Ali Jaffer	Camerman	Geo TV	2015	08-Sep-15	Sindh	Karachi	Shot dead	Yes	14	TV
Mahmood Syal	Freelance photo journalist	ATV	2007	28-Apr-07	Kuber Pakhtunkhwa	Chaman	Suicide blast	No	3	TV
Fazal Wahid	Freelance reporter	Daily Aard	2009	23-Jan-09	Kuber Pakhtunkhwa	Swat	Shot dead	Yes	5	Newspaper
Nasir Ali Rajul	HR Admin	Bal TV	2015	16-Jul-15	Sindh	Karachi	Shot dead	Yes	14	TV
Imdad Memon	Press Club Staffer	Press Club Staffer	2009	22-Oct-09	Kuber Pakhtunkhwa	Peshawar	Suicide blast	No	1	Press Club Staffer
Muhammad Syal	Reporter	Daily Nisr	2007	04-Jul-07	ICT	Islamabad	Killed in Crossfire	No	3	Newspaper
Muhammad Syal	Reporter	Daily Nisr	2007	18-Oct-07	Sindh	Karachi	Shot dead	No	14	Newspaper

Fig 8.3 – Original dataset (Causalities of Media Workers)

This particular dataset is what we commonly refer to as ‘record listing’, which means it has a (vertical) listing of records in each row. In order to count the total number of records, we simply need to see the ‘index’ given on the left side of the screen (as shown in figure 8.4).

The total number of records here is 112 (as shown in figure 8.5) but the number includes one title row as well (as shown in figure 8.6) which usually is the row number one and contains titles for each column. For instance, name, design (designation), org (organization) in this case.

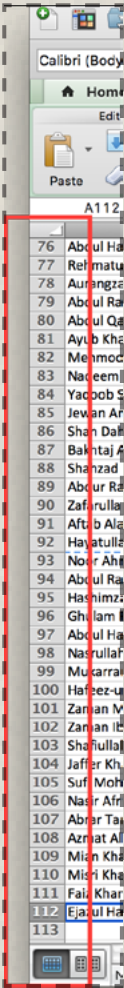


Fig 8.4 - Index in Microsoft Excel

Index	Name	Design	Org	Year	Reported Date	Province	City	Type	Targeted	Killed	Medium
76	Abdul Ha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
77	Rehmatu	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
78	Aurangza	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
79	Abdul Ra	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
80	Abdul Qa	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
81	Ayub Kha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
82	Mehmud	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
83	Nadeem	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
84	Yacoub S	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
85	Jewan Ar	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
86	Shan Dal	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
87	Bakhtaj A	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
88	Shanzad	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
89	Abdul Ra	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
90	Zafarulla	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
91	Aftab Ala	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
92	Hayatulla	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
93	Noor Ah	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
94	Abdul Ra	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
95	Hasnizam	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
96	Ghulam I	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
97	Abdul Ha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
98	Nasrulla	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
99	Muhammad	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
100	Harbez-u	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
101	Zaman M	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
102	Zaman M	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
103	Shahulla	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
104	Jaffar Kh	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
105	Sufi Moh	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
106	Nasir Afr	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
107	Abrar Ta	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
108	Azmat Al	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
109	Mian Kha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
110	Misri Kha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
111	Faiyaz Kha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
112	Ejazul Ha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
113											

Fig 8.5 - Total number of records including title row

Index	Name	Design	Org	Year	Reported Date	Province	City	Type	Targeted	Killed	Medium
1	Abdul Ha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
2	Rehmatu	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
3	Aurangza	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
4	Abdul Ra	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
5	Abdul Qa	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
6	Ayub Kha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
7	Mehmud	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
8	Nadeem	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
9	Yacoub S	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
10	Jewan Ar	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
11	Shan Dal	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
12	Bakhtaj A	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
13	Shanzad	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
14	Abdul Ra	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
15	Zafarulla	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
16	Aftab Ala	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
17	Hayatulla	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
18	Noor Ah	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
19	Abdul Ra	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
20	Hasnizam	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
21	Ghulam I	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
22	Abdul Ha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
23	Nasrulla	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
24	Muhammad	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
25	Harbez-u	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
26	Zaman M	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
27	Zaman M	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
28	Shahulla	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
29	Jaffar Kh	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
30	Sufi Moh	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
31	Nasir Afr	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
32	Abrar Ta	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
33	Azmat Al	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
34	Mian Kha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
35	Misri Kha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
36	Faiyaz Kha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		
37	Ejazul Ha	Reporter	Daily Smith	2010	10-May-10	Sindh	Shafa	Tortured to death	Yes		

Fig 8.6 - The title row

So according to the number of listed records in the index (112), minus the title row (1), we can say the total number of actual records are 111, or in other words, the total number of journalists who died in the line of duty are 111.

Question 2: What is, according to our data, the deadliest role (reporter, cameraman etc.) for a media worker?

To answer this question, we will have to move a step ahead from reading values from the default index. We can use multiple Microsoft Excel functions to get our answer, but since we are only doing a visual analysis, let us try and use the most basic functions only - in this case ‘sort’. You can apply sort function on any ‘one’ column of the dataset. To apply sort, select the desired column and press the sort icon from the toolbar above (as shown in figure 8.7a).

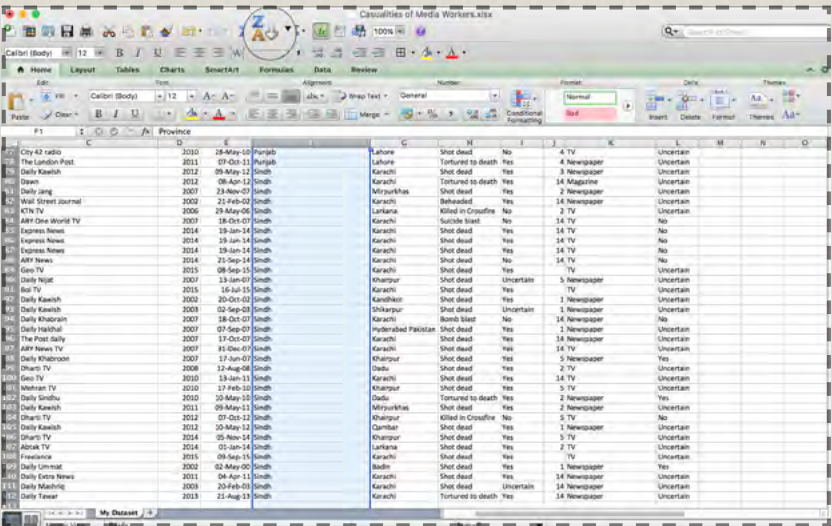


Fig 8.7a - Applying 'sort' function

Sorting function allows the user to sort complete dataset with reference to one column in ascending or descending order. For instance, if you apply the 'sort' function on the 'name' column, all of the records (in the name column) will be sorted from either A-Z or Z-A.

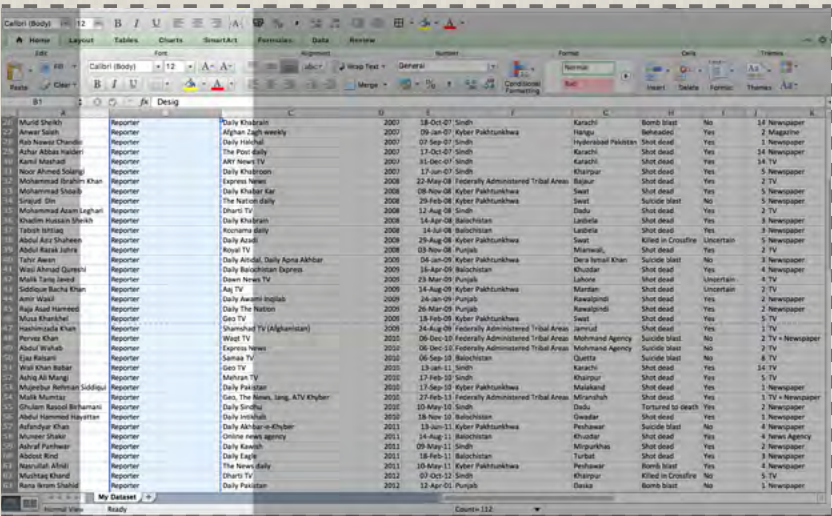


Fig 8.7b - Sorting function applied on 'design' column

Now that we have an idea of this data, let us try and use a little more sophisticated technique to answer the same question using the 'filter' option. Filter, as the name suggest, allows the user to filter the dataset using a single criterion or a set of criteria. Filter can be applied to one column or the whole sheet depending on our requirement. Let us proceed with the former.

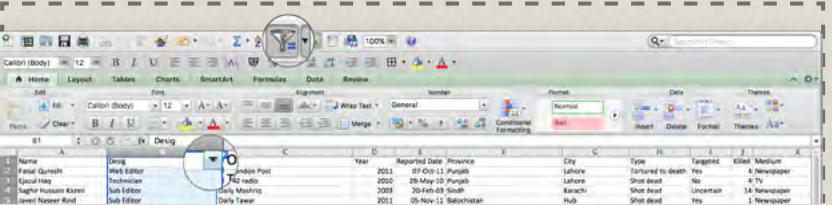


Fig 8.8 - The filter function

To apply a filter, select the desired row ('design' in this case) and press the filter icon in the toolbar above (as shown in figure 8.8)

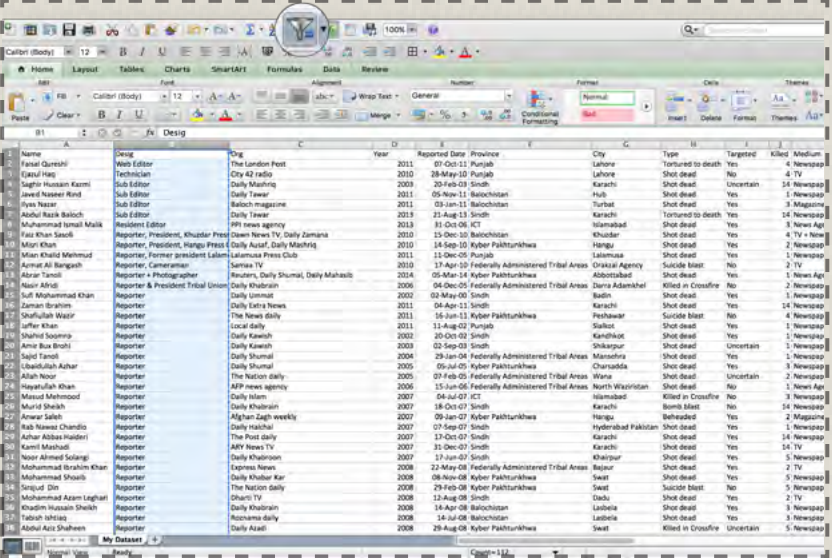


Fig 8.10 - Applying 'filter'

Pressing the dropdown menu button (highlighted in figure 8.10) will bring down a menu using which you can select the criteria of filtering (as shown in figure 8.11). Notice the ‘contains’ field, using which you can input the filtering criteria, in our case ‘reporter’.

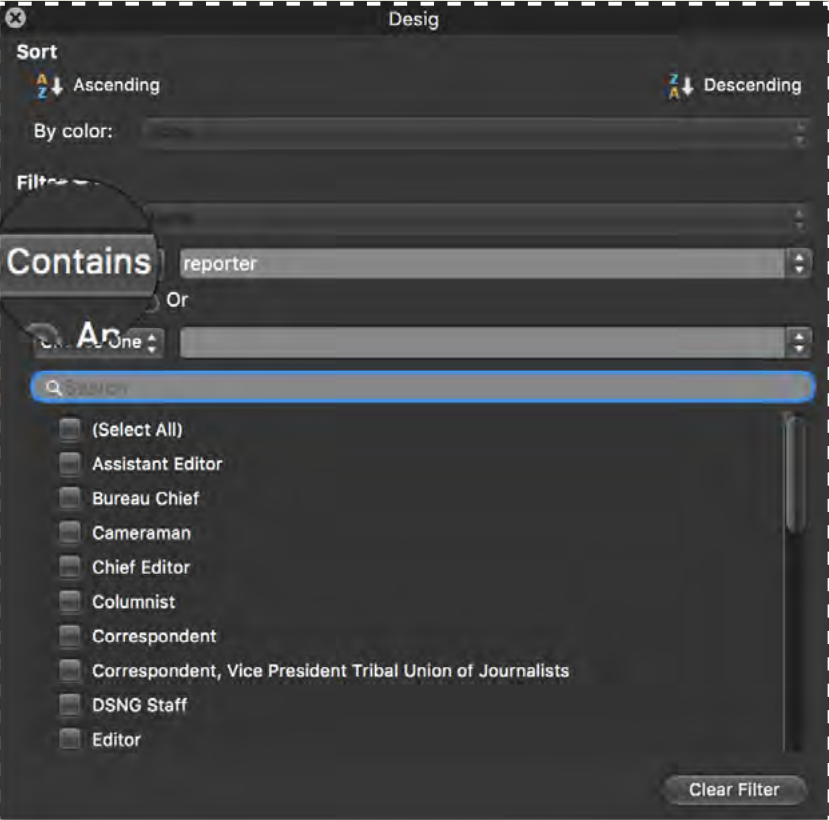


Fig 8.11 – Applying ‘filter’

Filtering the dataset using this criterion will pullout records with the word ‘reporter’ in the ‘desig’ column. Notice (in figure 8.12) that all records, which contain the word ‘reporter’ including ‘reporter + cameraman’ and ‘reporter + photographer’ are shown.

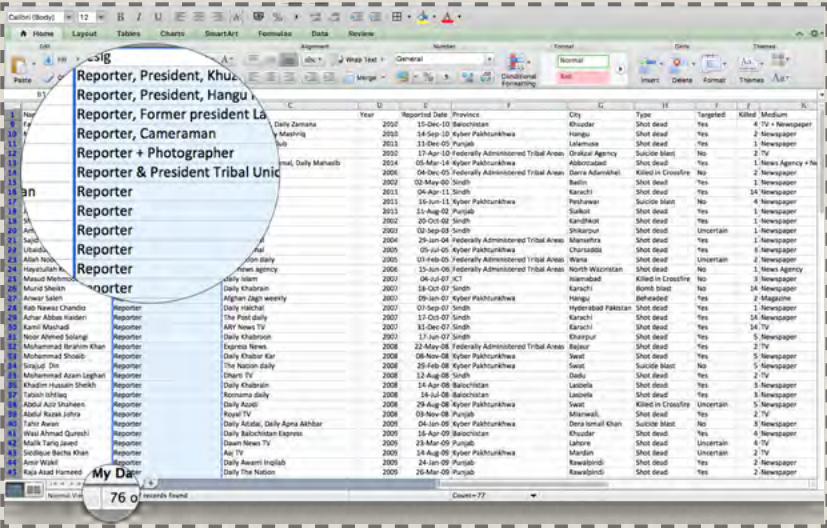


Fig 8.12 – Applying filter

If we observe closely, the total number of records is also shown in the status bar – which is 76. So, we can easily conclude that those who died in the line of duty, out of 111, 76 had the role of reporters.

Assessing the data, we can conclude that reporter is the most dangerous role for a media worker.

Question 3: Which of the provinces has proven to be deadliest for media workers?

Similar to questions one and two, we can use ‘sort’ and ‘filter’ functions to determine the answer. Using the ‘sort’ on the column ‘province’ (as shown in figure 8.13) will allow us to visually detect most recurring names (of provinces).

For instance, in figure 8.13, we can see an obvious majority of ‘Sindh’.

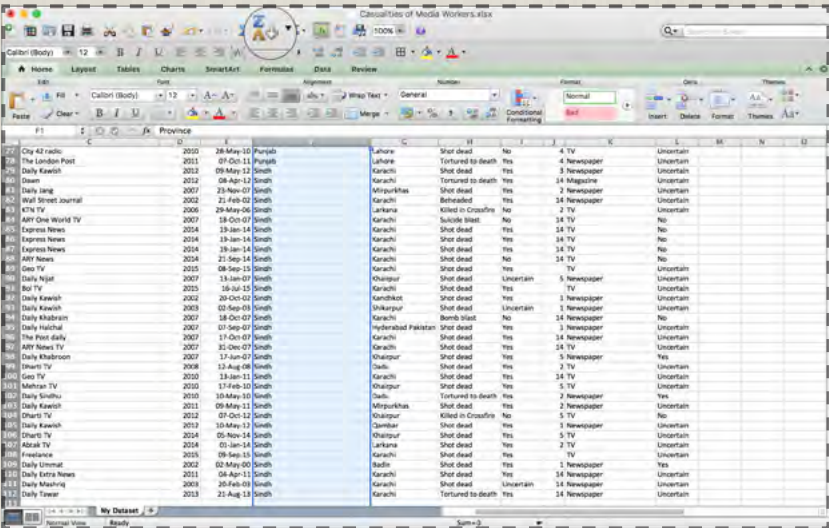


Fig 8.13 – Determining the deadliest province using ‘sort’

We can also determine this by using the ‘filter’ option. As in the questions above, to use filter, select the row you need to filter, in this case ‘province’ and press the filter icon in the toolbar above (as shown in figure 8.8). A dropdown menu button should appear next to the column you are trying to filter (as shown in figure 8.9). When you press the dropdown menu button, a dialogue box appears using which you can select or input the filtering criteria (as shown in figure 8.14).

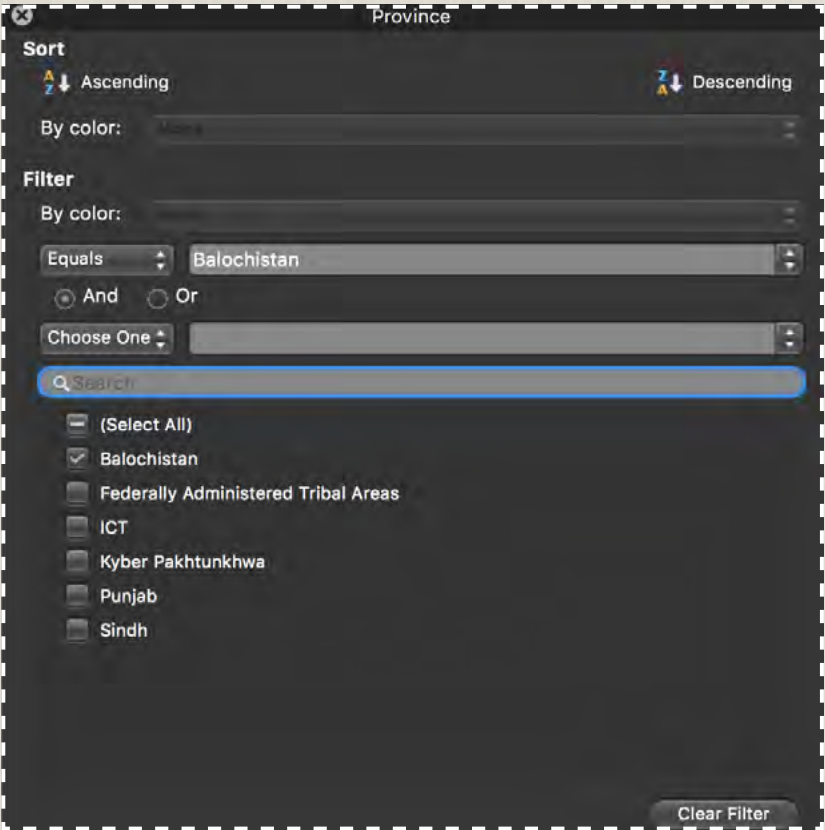


Fig 8.14 – Determining the deadliest province using ‘filter’

Notice that the dialogue box already has grouped entries from the ‘province’ column. Selecting one of them will only display the rows where there is a matching province name. For instance, as shown in figure 8.14, if we select ‘Balochistan’ see what happens (as shown in figure 8.15).

	A	B	C	D	E	F	G	H	I	J	K
Zafarullah Jatak	Reporter	Daily Inshad	2015	Balochistan	Jaffarabad	Shot dead	Yes	Newspaper	Uncertain		
Uzair Taseel	Reporter	Daily Inshad	2004	Federally Ad	Mawatha	Shot dead	Yes	Newspaper	Uncertain		
Muhammad Bux Khan	Reporter	Express News	2008	Federally Ad	Bejpur	Shot dead	Yes	TV	No		
Muhammad Khan	Reporter	Shimshad TV (Islamabad)	2008	Federally Ad	Lamard	Shot dead	Yes	TV	Yes		
Muhammad Khan	Reporter	Gaz, The News, Lang, ATY Khyber	2010	Federally Ad	Miranshah	Shot dead	Yes	TV + Newspaper	Uncertain		
Muhammad Ismail Malik	Resident Editor	PFI news agency	2013	ICT	Islamabad	Shot dead	Yes	News Agency	Uncertain		
Faiz Mahab	Investment reporter	Daily Ahsan	2003	Kyber Pakht	Swat	Shot dead	Yes	Newspaper	Yes		
Ustadullah Aulter	Reporter	Daily Inshad	2008	Kyber Pakht	Charsadda	Shot dead	Yes	Newspaper	Uncertain		
Masood Iqbal	Online news agency	2004	Kyber Pakht	Dera Ismail	Shot dead	Yes	Yes	News Agency	Yes		
Anwar Suleh	Reporter	Alghar Jagh weekly	2007	Kyber Pakht	Nangra	Beheaded	Yes	Magazine	Uncertain		
Muhammad Shoaib	Reporter	Daily Khatam Kar	2008	Kyber Pakht	Swat	Shot dead	Yes	Newspaper	No		
Muhammad Shoaib	Reporter	Gaz TV	2008	Kyber Pakht	Swat	Shot dead	Yes	TV	Uncertain		
Muhammad Shoaib	Reporter	Daily Ahsan	2010	Kyber Pakht	Nangra	Shot dead	Yes	Newspaper	Yes		
Muhammad Shoaib	Reporter	Daily Inshad	2010	Kyber Pakht	Mardan	Shot dead	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	The News daily	2011	Kyber Pakht	Punahwar	Bomb blast	Yes	Newspaper	Yes		
Muhammad Shoaib	Reporter	Dunya News TV, Voice of America	2012	Kyber Pakht	Charsadda	Shot dead	Yes	TV + Radio	Yes		
Muhammad Shoaib	Reporter	Kash Times	2013	Kyber Pakht	Karai	Shot dead	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Weekly Munawar-e-Aam	2014	Kyber Pakht	Mardan	Beheaded	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Radical, Daily Inshad, Daily Mahab	2014	Kyber Pakht	Abdullahabad	Shot dead	Yes	News Agency + Newspaper	Yes		
Muhammad Shoaib	Reporter	No Base, Daily Inshad	2015	Kyber Pakht	Tara	Shot dead	Yes	TV + Newspaper	Yes		
Muhammad Shoaib	Reporter	Neo TV, Daily Ahsan	2015	Kyber Pakht	Kuhai	Shot dead	Yes	TV + Newspaper	Yes		
Muhammad Shoaib	Reporter	Lahore daily	2011	Punjab	Gulistan	Shot dead	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Lahore Press Club	2011	Punjab	Lahore	Shot dead	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Royal TV	2008	Punjab	Mianwali	Shot dead	Yes	TV	Yes		
Muhammad Shoaib	Chief Editor	Daily London News	2008	Punjab	Lahore	Shot dead	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Daily Kashmiri Inqilab	2009	Punjab	Rawalpindi	Shot dead	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Daily The Nation	2009	Punjab	Rawalpindi	Shot dead	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Asia Times Online	2011	Punjab	Mandi Bahau	Tortured to death	Yes	Newspaper	Yes		
Muhammad Shoaib	Web Editor	The London Post	2011	Punjab	Lahore	Tortured to death	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Samaa	2011	Punjab	Mianwali	Shot dead	Yes	TV	Uncertain		
Muhammad Shoaib	Reporter	Dunya daily	2014	Punjab	hafizabad	Shot dead	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Express News TV and Express daily	2014	Punjab	hafizabad	Shot dead	Yes	TV + Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Daily Inshad	2002	Sindh	Karai	Shot dead	Yes	Newspaper	Yes		
Muhammad Shoaib	Reporter	Wall Street Journal	2002	Sindh	Karachi	Beheaded	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Daily Karachi	2002	Sindh	Karachi	Shot dead	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	Daily Khairpur	2007	Sindh	Khairpur	Shot dead	Yes	Newspaper	Yes		
Muhammad Shoaib	Reporter	Daily Inshad	2007	Sindh	Hyderabad F	Shot dead	Yes	Newspaper	Uncertain		
Muhammad Shoaib	Reporter	The Post daily	2007	Sindh	Karachi	Shot dead	Yes	Newspaper	Uncertain		

Fig 8.17 - Determining the number of targeted media workers using 'sort'

As you can see (from figure 8.17) the 'yes' has the majority. But we still do not have a number, as to how many were actually targeted. Let us try and get it using the 'filter' option.

Select the column titled 'targeted' and press the filter button on the toolbar (as shown in figure 8.10). A dialogue box will appear (as shown in figure 8.18), using which you can select the criteria. In our case, since we want to see the number of those who were targeted, we will select 'yes'.

Targeted

Sort

Ascending Descending

By color:

Filter

By color:

Choose One

Search

(Select All)

No

Uncertain

Yes

Clear Filter

Observe in figure 8.19, that the records displayed now are of only those who were targeted, along with the number in the bottom left corner.

SECTION FOUR FOR THE EYE OF THE BEHOLDER



CHAPTER NINE

BASICS OF DATA VISUALISATION

ONCE the numbers have told you the story, you need to tell it to your readers. But, your readers might not be interested in looking at the intimidating spreadsheets or the complex pivot tables. No, to connect the dots for your readers, you need to present the data in an interesting and visually appealing manner. In this chapter, we will discuss data visualization and the conceptual considerations that need to be made during the visualization process. We will look at the basic considerations you need to make when deciding on the form of visualization and give you a brief tutorial about a visualization tool, Infogram.

The six rules of visualization

1. Know your story

This about what you want to show your audience. Choose the message that you really want the reader to take, rather than simply visualizing every number you have access to. Unless you know what you want the infographic or data visual to say, your audience might be confused by the data.

2. Take time to plan

When you are working with complex datasets, there can be various visualization possibilities. You could make pie charts, bar graphs, scatter plots. You could make a word cloud

or map your data. What will be most effective, most striking and most likely to help your storytelling is something you will need time deciding. Explore different possibilities of visualization with your data. Do not settle - think about different visualization combinations. You could add a map and a word cloud. You could add multiple type of charts or infographics. Do not just go with your first instinct. Take time to plan and make sure that the visualizations do justice to your data and to your story.

3. Think Visually

This is perhaps one of the most important points of visualizations. Think about how you, yourself will react if the graphic/infographic you are about to make was presented to you. The human brain is geared to take in information visually. Our working memory is limited, and visuals have the power to immediately grab attention. If you plan your visual well (both in terms of clarity and aesthetics), it might just become the key take away that you readers remember.

4. Consistency is key

This is a simple rule - when working on one project, make sure that your design is consistent throughout. Do not use multiple fonts. Do not jump between various shades of colors or change proportions of graphs in

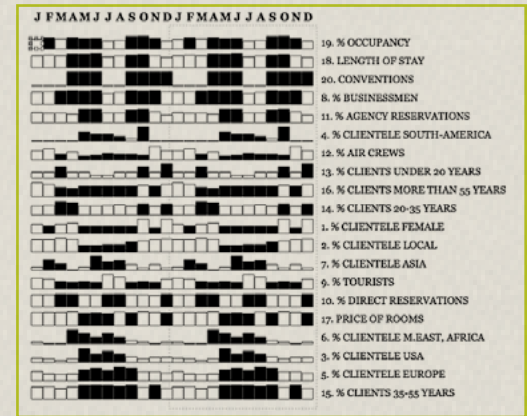
different sections of the story. In your design, consistency is the key that will hold your piece together. In addition, a piece that includes different fonts, different families of colors and an inconsistent design can look unprofessional and unedited.

5. Be Clear

Think about the average news consumer. Most of them have a thousand other content options and not a lot of time. To get your story out and to get your message across quickly, you need to be clear. This means that your design should be simple, the choice of graphics clean and self-explanatory and the caption enlightening.

6. Keep it simple

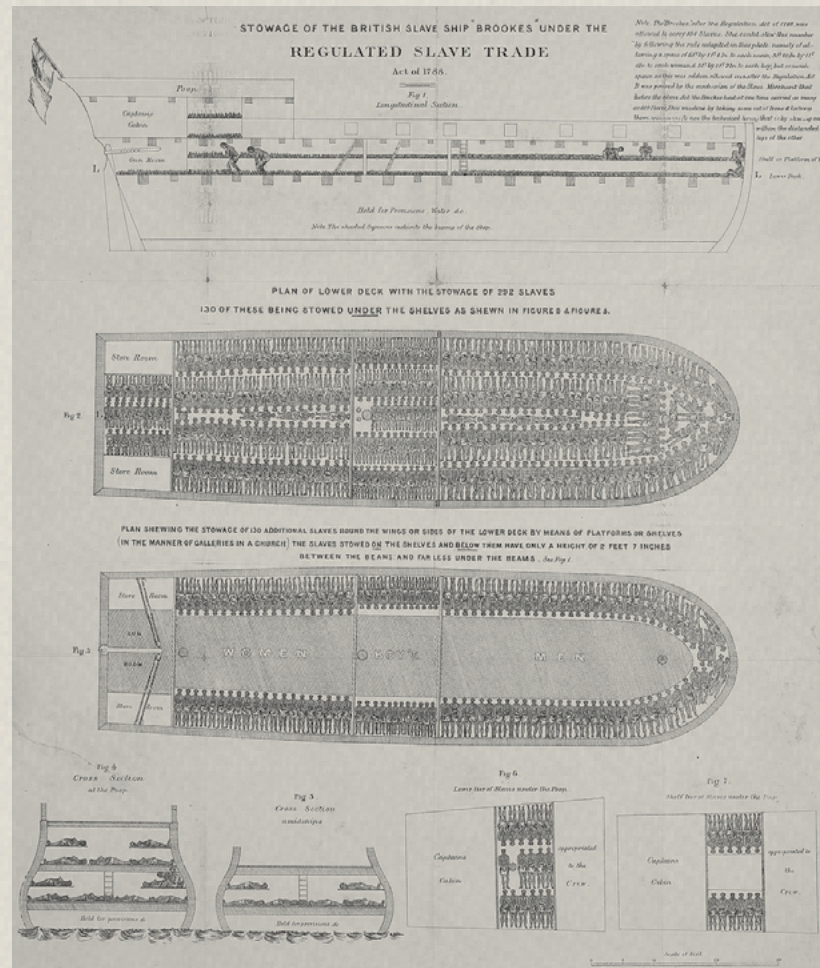
Nowadays, it is fairly simple to get access to and use simple tools to generate really complex, fancy looking graphics. In addition, when you work with a dataset that records hundreds or thousands of values and incorporates multiple variables, it is very easy to be tempted to make really complex graphics. After all, the data is complex. But, guard yourself from this temptation or you could end up like this!



A graphic on the tool Lyra – Can you understand what it says?

The lesson is, the simpler your visualization is, the better it will be understood.

The rules we have outlines are simply to help you formulate your approach towards data visualization. The impact of visual information cannot be over emphasized. Before moving on to Infogram, here is a historic visual image from 1789, that presented historical data about slave trade, that was at that time a legal, regulated form of trade in England.

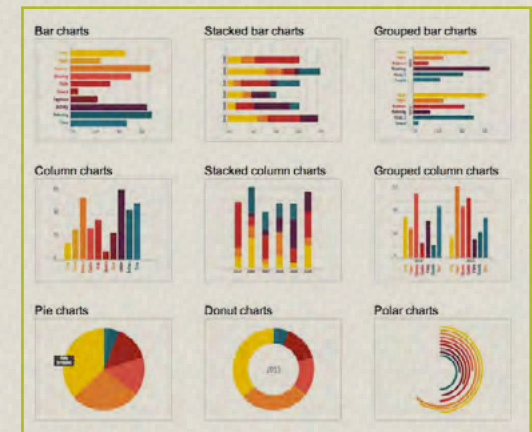


A poster showing how people were packed like cargo into the ship, forced to lie on their sides and unable to stand for up to six months at a time.

This poster later became a key asset for groups that were advocating against slavery. As data journalists, you might come across figures that shock and awe you. Creative and thoughtful visualization would help you transfer that sense of shock and awe to your reader, adding to the impact of your story. With that, let us look at a visualization tool that you can use to create beautiful graphics and infographics with minimal effort.

Infogram

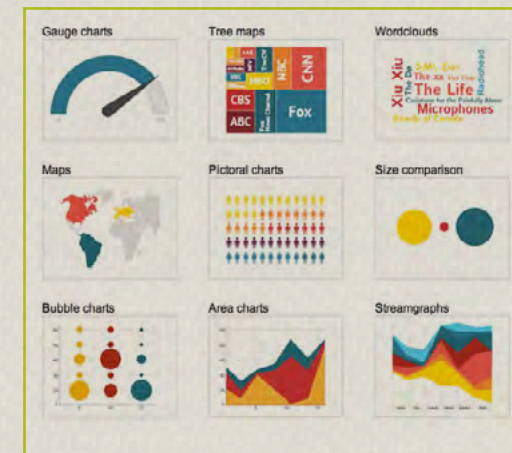
Infogram is a free web based application through which you can create online graphs, graphics and infographics. This is one of the simplest tools available online. You can use it to create different types of charts, graphs, clouds, maps and infographics.

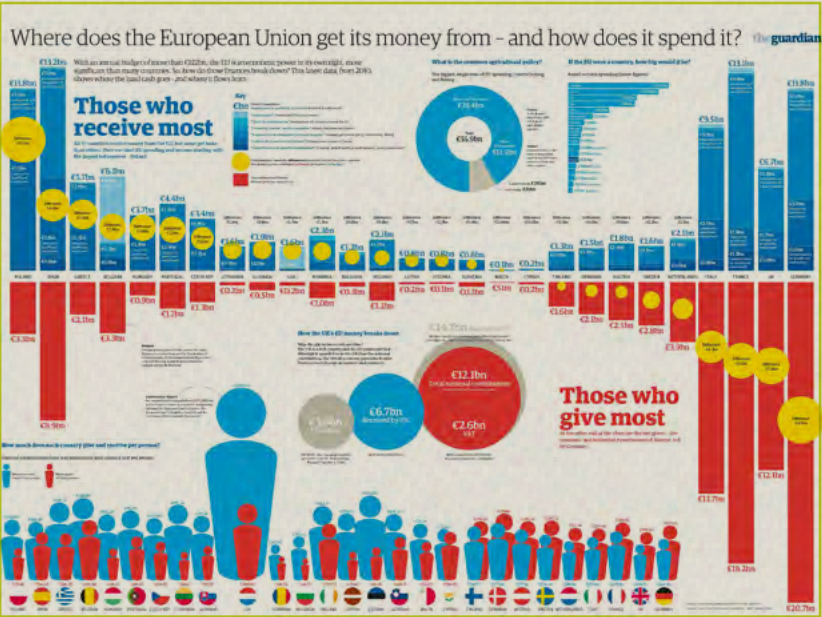


Some of the chart types you can create on Infogram

A detailed guide to which charts to use for what kind of data, prepared by Infogram is attached as Annex 2.

You can directly mix and match different types of graphs on a single canvas, which allows you to tell different angles of the story through the same image.





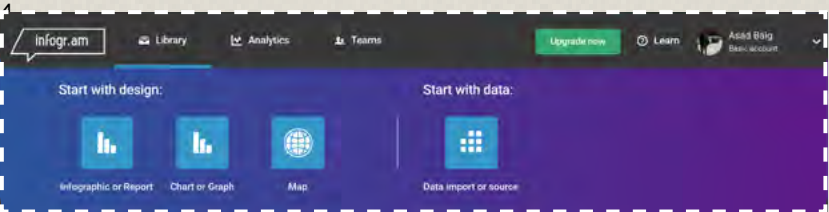
An infographic in The Guardian

Using Infogram you can make similar, professional looking infographics in a few minutes, without having to learn a whole lot of new techniques. Infogram also has a mapping function, however, the default map does not include Jammu Kashmir within Pakistan, which might be problematic in the national context.

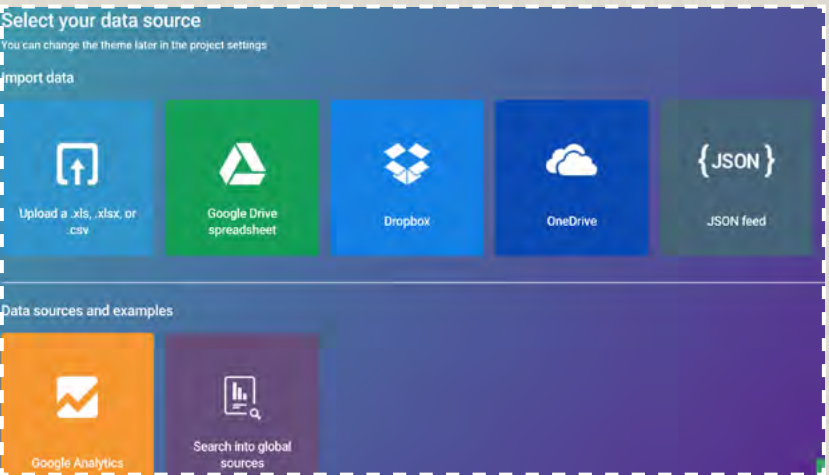
You can also share your infographic directly on social media and other online platforms. If you are using a paid version you can also download the infographics in multiple, high resolution formats. The free version unfortunately doesn't allow downloads, but is great for creating content that you have to use online.

Working with Infogram

- 1. Go to the website infogr.am
- 2. Log in using your Facebook, Google Plus, Twitter or LinkedIn account or create a new account
- 3. Choose the type of visualization you want to make. You can choose between chart or graph, infographic/report or map.



- 5. As you can see, you can also simply start by uploading your dataset you have and choose the style later. Infogram allows you to upload data from spreadsheets i.e. xls or xlsx formats, csv format.



You can upload data on infogr.am through any of these sources

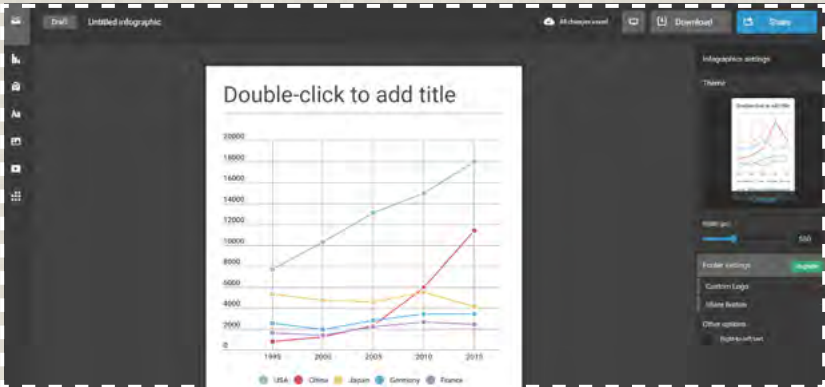
If you are trying to create infographics, Infogram offers you some free themes that have preset design elements. You can edit everything in a theme as well.



Free themes for infographics on Infogram

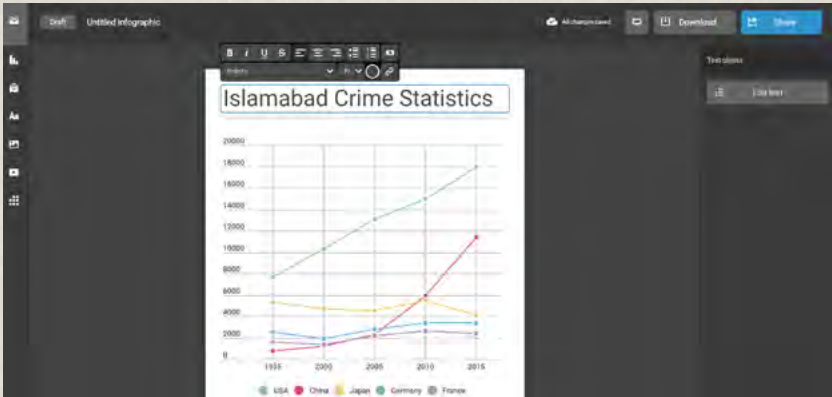
If you get a subscription to the site, you can find a larger range of paid themes.

5. Choose the theme you like and double click to open the Infogram editor.

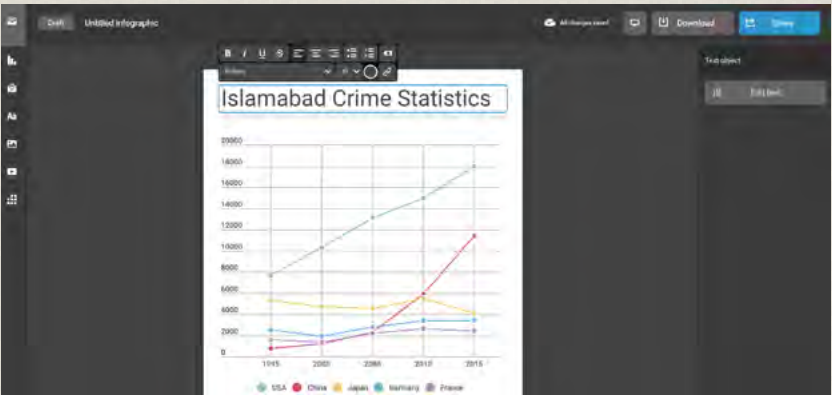


For this manual, we will be using statistics by Islamabad Police to create an infographic.

When you double click the title, and add the desired text, you will also see the text editor. You can use this text editor to customize the text in whichever design you like.



If you want to add more text, for example a subtitle under the main title, you can simply click the Aa on the left side bar.



You can choose the style of text, that goes best with the content you want to add. You can also use the rich text editor, that appears when you double click a text in the infographic to change fonts, modify font size and color, make it bold, italic, underline or strike through and align the direction and placement of the text.

Selecting the ideal text style for your infographic

Title is used to describe the topic of your project.

Subtitles are smaller in size and usually used to name each chart in your project.

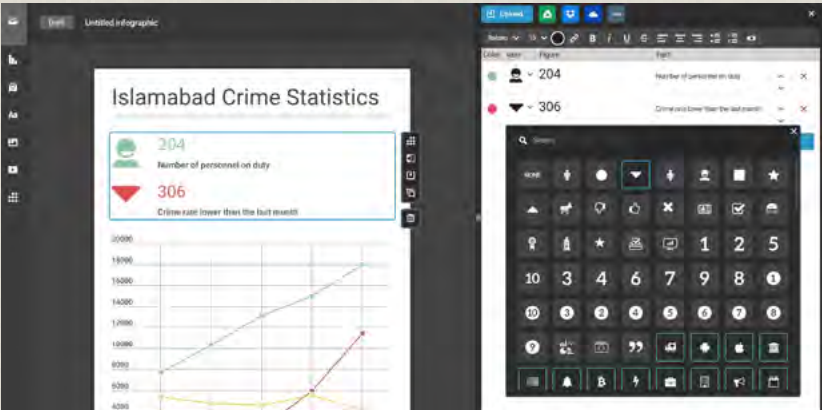
The quote is good for emphasising contextual information about your data. It has two lines of text that can be edited. The quotation is styled according to the colour theme you are using.

Body text is smaller than the subtitle text and is used to provide a lengthier explanation of your project or chart. This object can give your chart the proper context.

Caption text is the smallest text size usually used to add the source of your data.

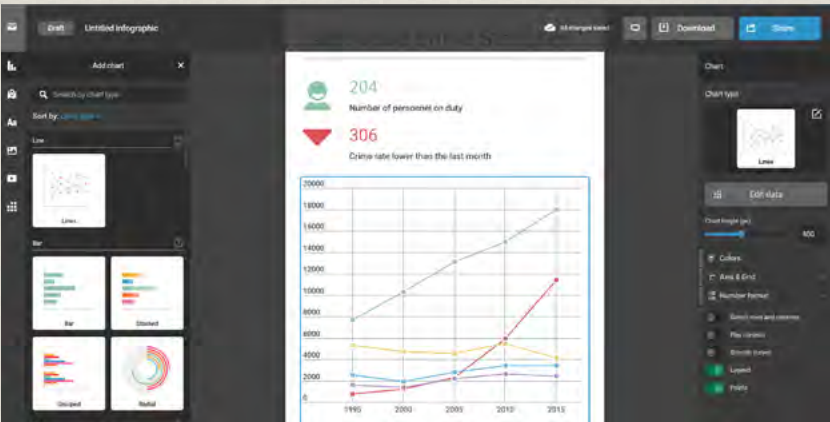
- Via Infogram

In the text pane, you will also find an option for facts and figures. Using this option, you can highlight the key statistics in your data.

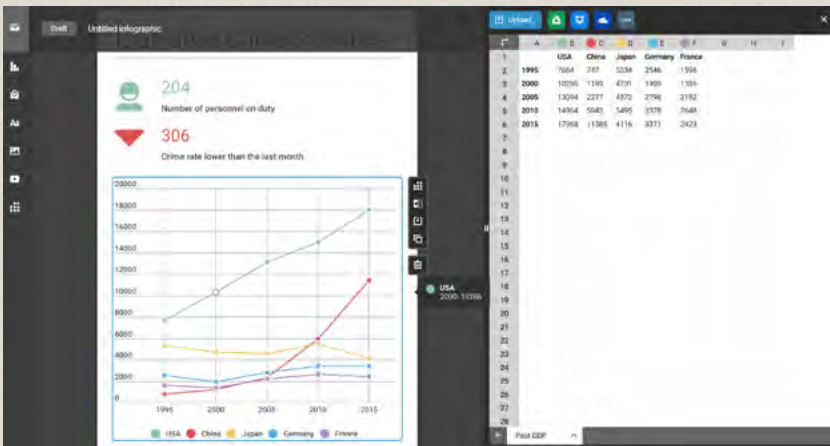


Once the facts and figure box has been inserted, double click it to open the editing panel that is shown above. Here you can edit icons and add your own text and figures.

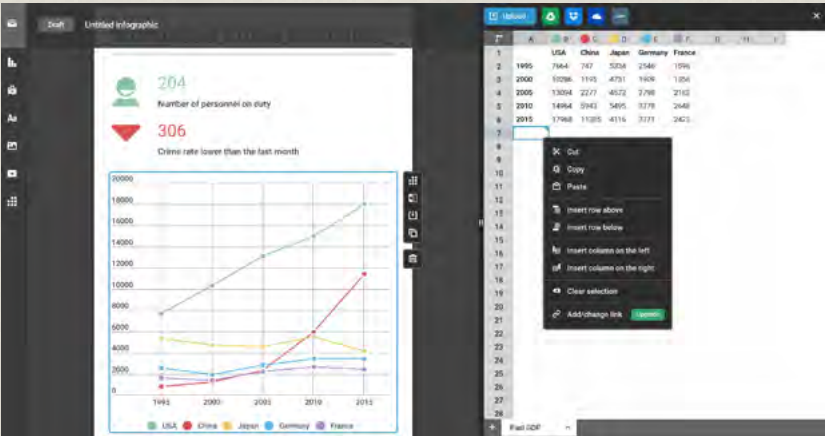
You can add a new graph by clicking on the bars at the top of the left- hand menu.



As you can see, the chart menu shows you all the different kinds of charts and graphs that you can use. Once you insert the chart in your infographic, the editing window will open on the right side. Using the editing window, you can manually add, or upload the data you want to visualize through the graph.



You can right click on any of the cells to get additional options to modify the dataset

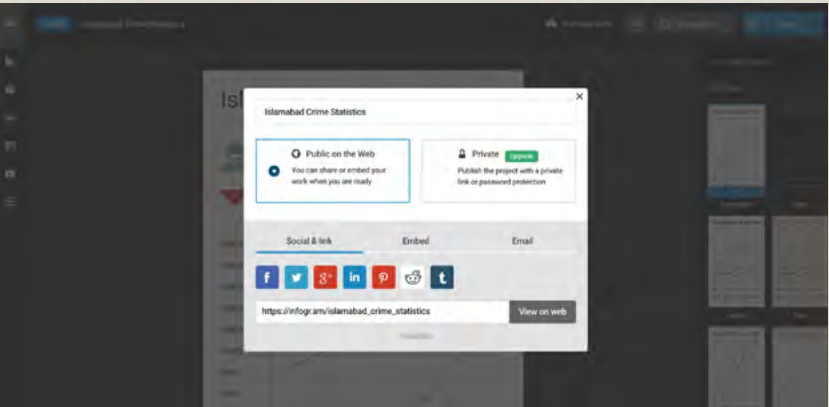


Using this you can insert rows and columns; delete rows and columns, cut, copy and paste data into the dataset. Clicking the circles shown on column headers will allow you to customize the colors of the lines as well.

Once the elements of your infographic have been finalized, you can preview how it looks by clicking the screen icon on the top right of the screen. If you are not happy with the look, you can also try different themes by double clicking the theme editor.



Once you find a theme and a look you like, simply click the blue Share button to share the infographic online.



If you have a subscribed version, you can also download it as PDF or PNG/ JPEG.

In this example, we have used a very simple example to demonstrate what you can do with Infogram. Once you start using the tool, you will quickly realize that there is potential to do so much more with this. If you are incorporating multimedia components within your story, Infogram also allows you to directly embed both video and images into your infographic.

CHAPTER TEN

DATA MAPPING

A MAP organizes wonder, said Ellen Meloy, an American nature writer. The maps we generate for journalistic stories are no less wondrous; they are amazing for the insight they offer and the clarity they bring to stories. In this chapter we will explore maps as a visualization method. We will explore some historical maps created by data analysis and introduce you to a basic mapping tool Fusion Tables, another Google tool that can be used for free and by people with beginner level technical skills.

The creation of maps to understand and tell stories is not a new phenomenon. In fact, one of the biggest medical breakthroughs owes thanks to a map created by the inquisitive British doctor John Snow.

A piece written for University of Delaware states that “Dr. John Snow is regarded as one of the founding fathers of modern epidemiology. As London suffered a series of cholera outbreaks during the mid-19th century, Snow theorized that cholera reproduced in the human body and was spread through contaminated water. This contradicted the prevailing theory that diseases were spread by ‘miasma’ in the air.”

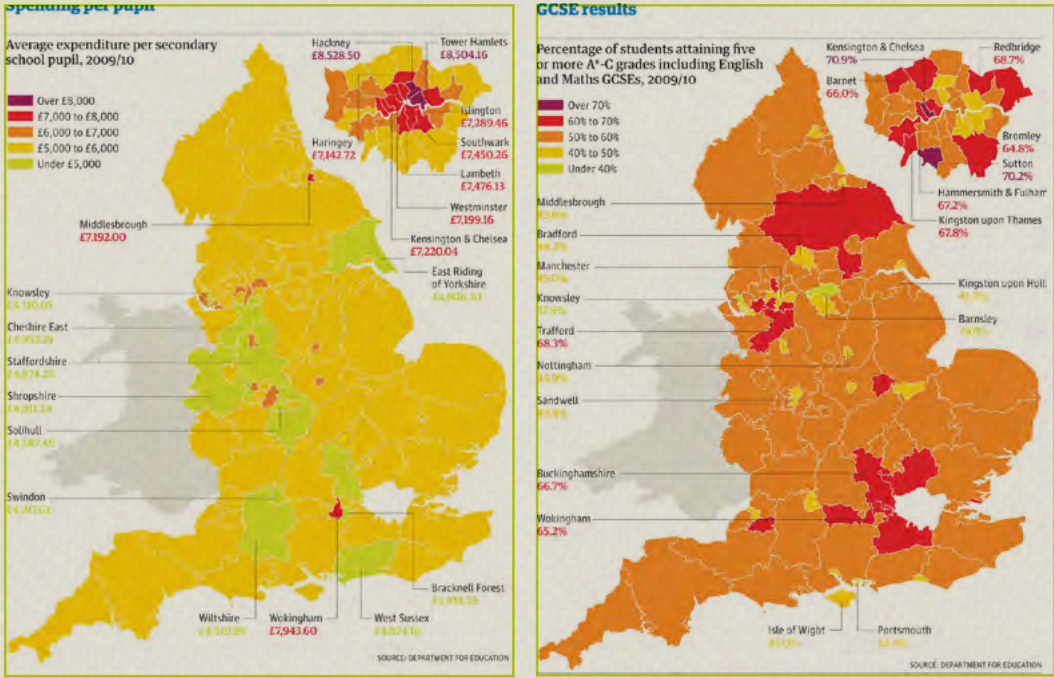
This means that Dr. Snow was the first one to theorize about the possibility of spread of disease by contaminated water. Behind his theory was a map. This map was painstakingly produced by Dr. Snow to plot deaths by cholera in London.



Dr. John Snow’s map of 1854 London cholera outbreak

You might be wondering what a map produced by a medical doctor, a scientist has to do with journalism. Well, journalists are social scientist and mapping data effectively can lead you to make discoveries that are quite significant and interesting.

Take a look at these two maps created by the Guardian Data Blog.



Guardian data blog maps data provided by UK department of Education

As you can see these maps demonstrate how education budget is spread throughout the country and also maps the results being produced by different schools across England. This is an excellent example of how maps can be used to visualize government preferences in budget allocation and its impact. You could find out the education or health budgets in different districts across Punjab or KP or other provinces and file an RTI request for results related data to create similar stories. Other maps could demonstrate average health statistics (like maternal mortality rates, average age etc.) across different districts in comparison to health budgets allocated and spent in the same. Putting this data on a map will give it a visual impact that would otherwise be missing from the story.

Here, you might be wondering how you, a journalist, with (most likely) a social science background and no training in graphic design or geo spatial mapping can create a map and put complex data on it! Have no fear, once again, it is Google to the rescue, with a free, easy to use tool that makes map mapping extremely easy. .

Google Fusion Tables

In this chapter, we will be focusing on a very useful tool created by Google – Google Fusion Tables, although still in a beta experimental form, provide a way to map complex data in less than ten clicks.

The maps copied above have been created by the Guardian using the same tool.

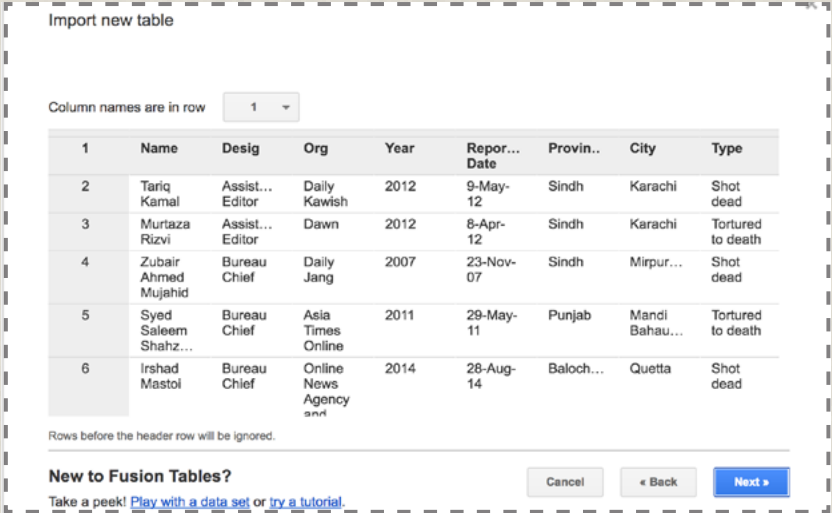
To use Fusion Tables, you must have a Google account. Here is how you make maps with Fusion Tables. You must have a dataset to map. Let us refer back to the dataset you have been using to practice all the other exercises in this manual, the Journalist safety data set. Here is how you can plot the journalists’ killings on Pakistan’s map.

Step 1 – Importing the Data

- 1. Sign in to your Google account and go to drive.google.com
- 2. Click the ‘New’ button
- 3. If you see an app called Fusion Tables, click it
- 4. If you do not see Fusion Tables, click ‘Connect more apps’, search for Fusion Tables and click ‘Connect’, click ‘OK’ when asked for confirmation.
- 5. Once Fusion Tables is connected you will get this screen



- 6. If you have the data saved on your computer, select ‘From this computer’ and then click ‘Choose file’. Attach the file (in this case the media workers casualty data). If you have it on your Google Drive, you can also directly add the link by selecting ‘Google spreadsheets’.
- 7. After your file name appears beside ‘Choose file’, click ‘Next’ Fusion Tables will upload your data and show you a preview.



- 8. Check that the data is formatted correctly and click ‘Next’, you will get the following screen.



9. On this screen add any details you would like to attribute to the dataset you are working with and click 'Finish'. Your dataset should appear like this:

Casualties of Media Workers

Imported at Sun May 28 14:47:50 PST 2017 from Casualties of Media Workers.xlsx
Exported at 02:52

File Edit Tools Help Rows 1 Cards 1 Map of Province

Filter No filters applied

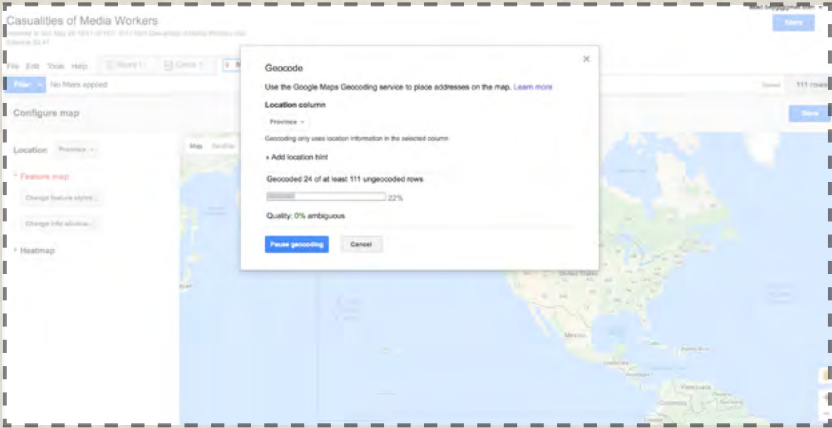
1-100 of 111

Name	Design	Org	Year	Reported Date	Province	City	Type	Targeted	Killed	Medium	Threatened
Tariq Kamal	Assistant Editor	Daily Kaash	2012	8-May-12	South	Karachi	Shot dead	Yes	3	Newspaper	Uncertain
Murtaza Rizvi	Assistant Editor	Dawn	2012	8-Apr-12	South	Karachi	Tortured in death	Yes	14	Magazine	Uncertain
Zubeir Ahmed Mujahid	Bureau Chief	Daily Jang	2007	23-Nov-07	South	Mirpurkhas	Shot dead	Yes	2	Newspaper	Uncertain
Syed Saleem Shahzad	Bureau Chief	Azad Times Online	2011	29-May-11	Punjab	Mirpurkhas	Tortured to death	Yes	1	Newspaper	Yes
Inshad Mastoi	Bureau Chief	Online News Agency and ARY News TV	2014	28-Aug-14	Kashmiristan	Quetta	Shot dead	Yes	8	News Agency + TV	Uncertain
Danish Pearl	Bureau Chief	Wall Street Journal	2002	21-Feb-02	South	Karachi	Beheaded	Yes	14	Newspaper	Uncertain
Mir Nawaz	Camerasman	AP news agency	2005	7-Feb-05	Kashmiristan	Wazir	Shot dead	No	2	News Agency	Uncertain
Javed Khan	Camerasman	DM Digital (UK based)	2006	4-Jul-07	CT	Islamabad	Killed in Crossfire	No	3	TV	No
Mune Ahmed Sangi	Camerasman	KTN TV	2006	29-May-06	South	Karachi	Killed in Crossfire	No	2	TV	Uncertain
Muhammad Arif	Camerasman	ARY One World	2007	18-Oct-07	South	Karachi	Suicide blast	No	14	TV	No

You have successfully imported the dataset that you want to map. If you look at the last screenshot (or the imported data on your own Fusion Tables), you will see that two of the columns have been highlighted in yellow. In our dataset these columns are the province and the city columns i.e. both columns that include data that can be geographically plotted on a map.

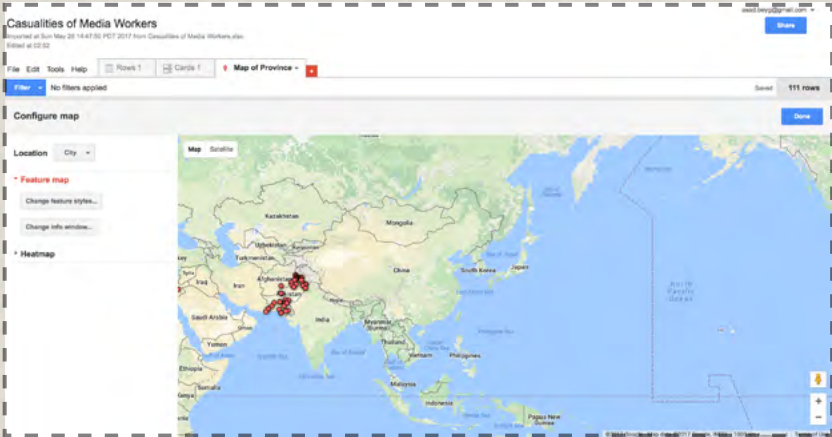
Map the data

1. Take another look at the imported data set (screenshot above). You will see a tab called 'Map of provinces'. The name of this tab will change in accordance with the location columns in each unique dataset. So, if the dataset we are using had a column called sectors and included the location according to sectors in Islamabad, the last tab will be called map of sectors. Similarly, if you had the location marked by longitude/latitude, you will get a tab titled 'Map of longitude/latitude. To map the data, simply click on this tab. You will get the following screen:

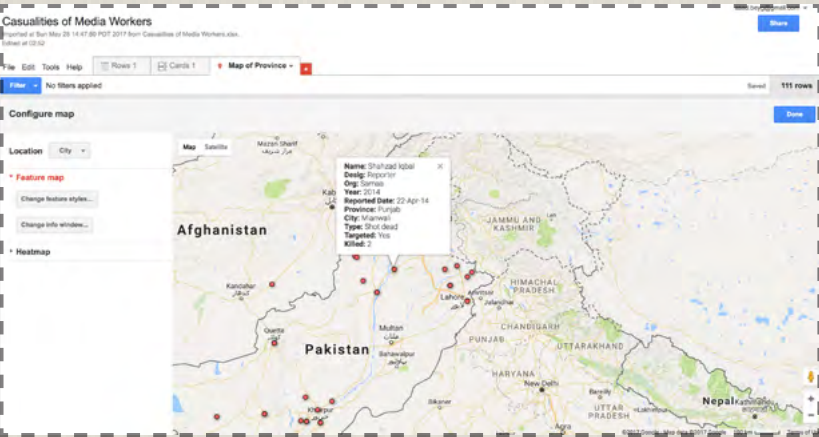


This means that Google is now using Google maps to plot or geocode your data on the world map. The geocoding can take time depending on the number of entries/plotting points your data has.

2. Now, since the dataset we are using has two different location columns i.e. province and city, we can choose which column Google uses to plot the data. Once the initial geocoding has finished, click the 'Location' tab on the left of the screen and choose 'City', as shown in the screenshot below:



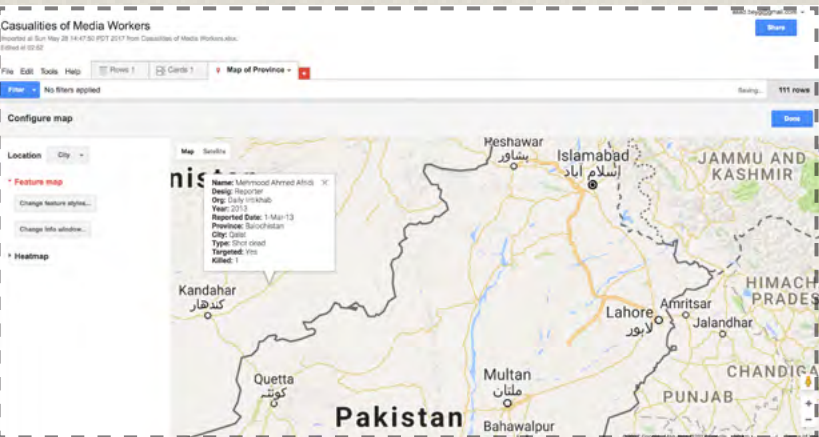
3. Here is a closer look at the map. When you click at any of the red pointers plotting the killings, you will get the related information.



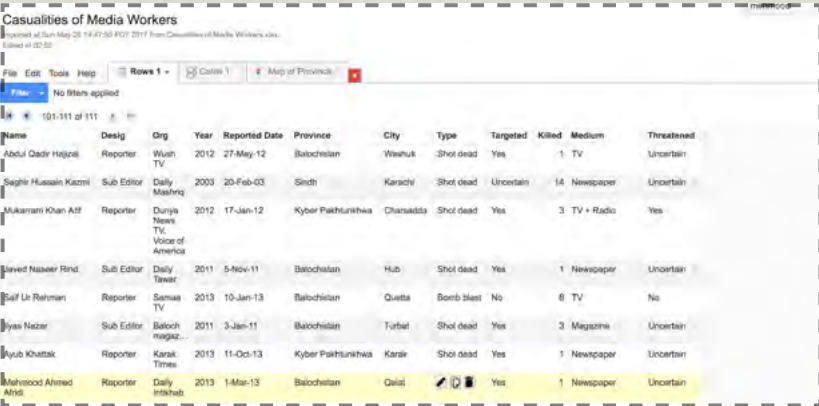
Correcting misplaced plots

As you can see, one of the pointers in the screenshot above, appears within Afghanistan’s boundaries. Since Fusion Tables is still an experimental tool and since there are similarly named cities and places in different countries, sometimes you will find that the mapping procedure misplaces a few entries. To correct these follow the following steps;

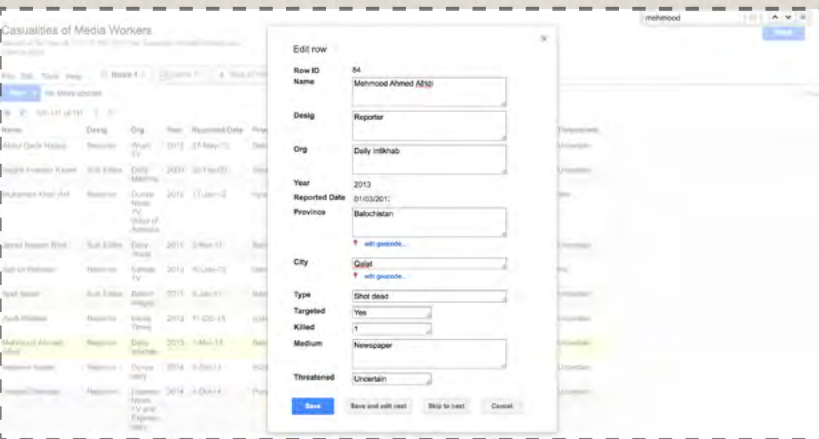
1. Click the misplaced pointer to find out which entry has been misplaced, in this case, it is Mehmood Ahmed Afridi from Qalat.



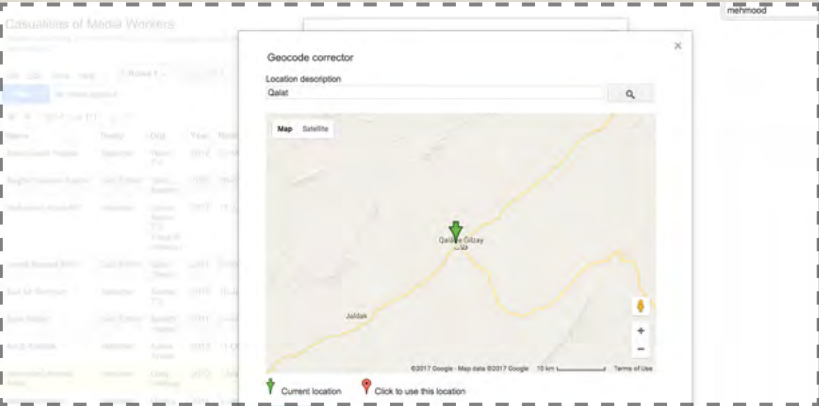
3. Go back to the ‘Rows 1’ tab, find the entry and click on the ‘Location’ tab. You will get this screen.



4. Double click the first icon i.e. the pen on the highlighted entry to open the editing tab.

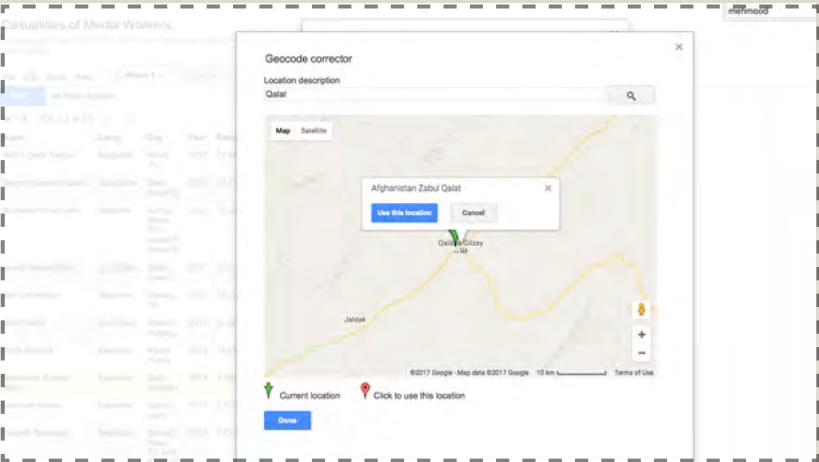


5. Click 'Edit Geocode' beneath the 'City' tab. If you are using another data-set, click 'Edit Geocode' beneath the 'Location' tab.



When this screen appears, add hints about the correct location in front of your original entry in the location description. Like in this example, we have added Balochistan, Pakistan as additional hints for correct location. After adding the hints, click the search icon in front of the description space.

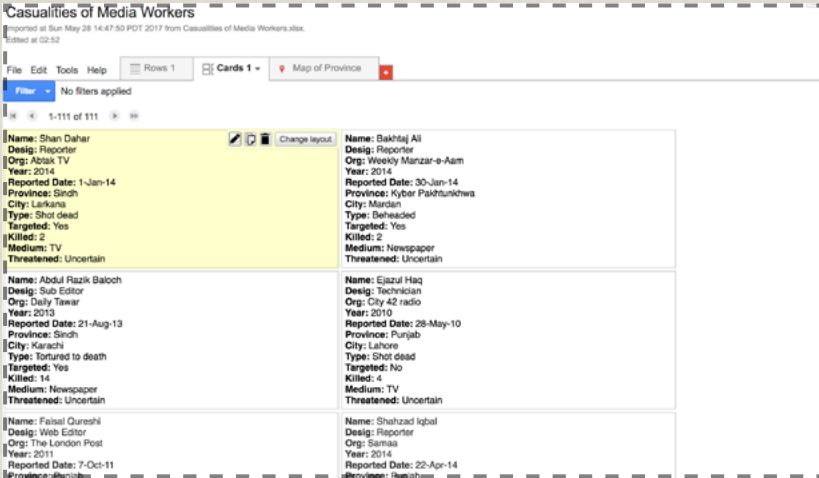
6. Google will search and place a red plotter on the new location. Click the red locator and you will get a prompt to confirm whether you want to edit the geocode to the new location. Click 'Use this location' and click 'Save'.



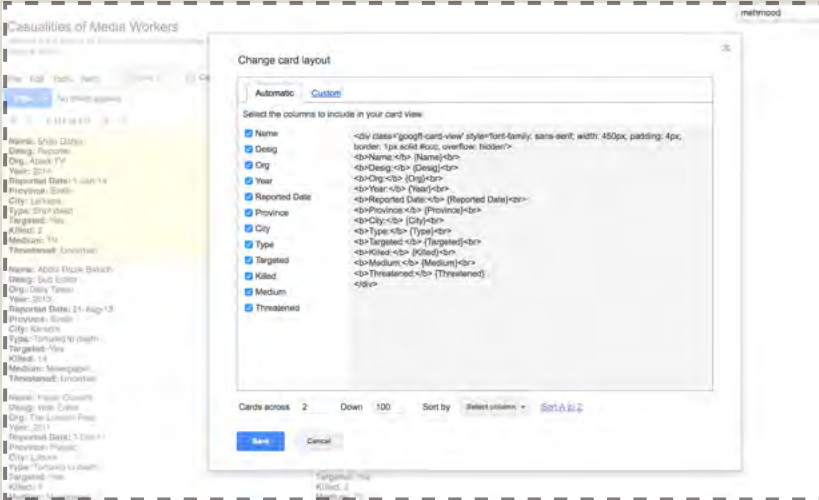
Repeat for all entries that have been wrongly plotted.

Editing labels

You can also change the way entry labels look when someone clicks on them. To edit labels, simply go to the 'Cards' tab, double click any entry and click 'Change Layout' as shown in the screenshot below.



You will get this screen when you click 'Change Layout':

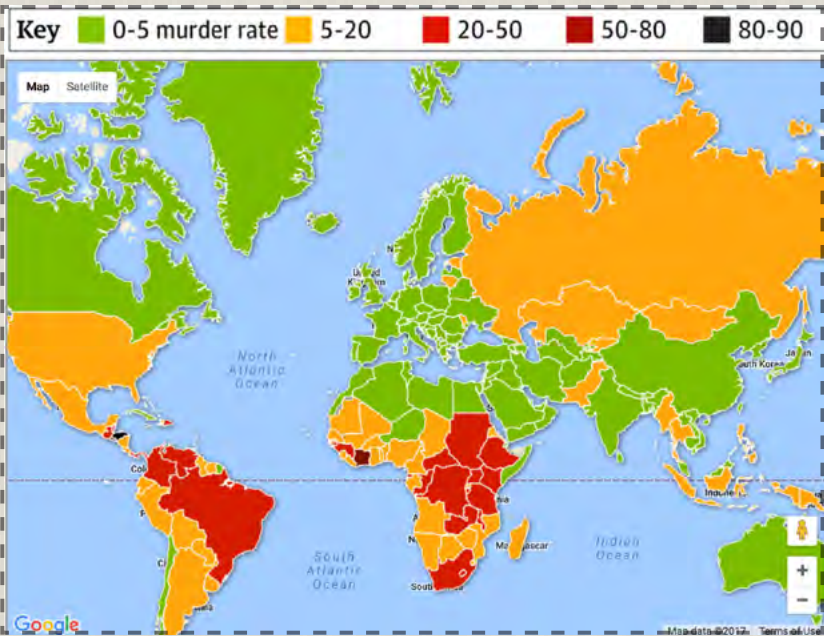


As you can see, the screen lists all the names of the columns in your original dataset. You can choose which parts of this data appear on the card that is displayed when anyone clicks a pointer on the map. Simply uncheck the fields that you do not want to appear on the label.

Some Inspiration

Here are some examples of how news outlets around the world are using Fusion Tables.

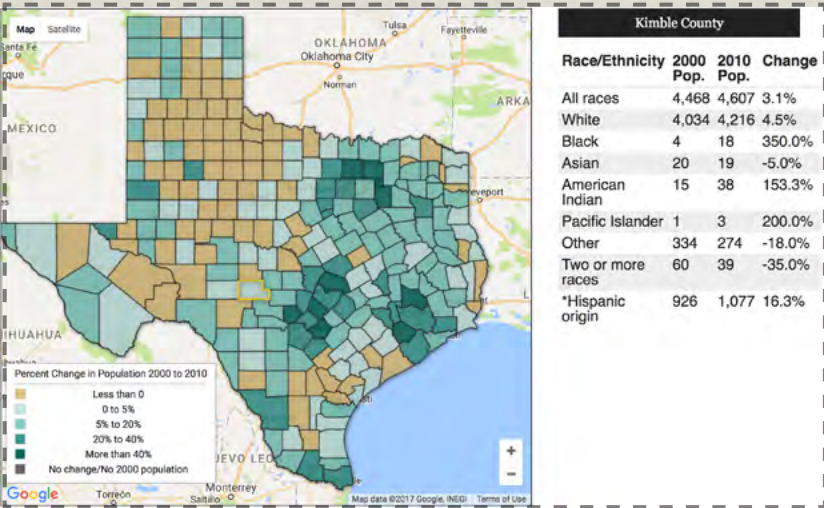
Murders around the world



This simple map was created by The Guardian Data Blog, mapping the rate of murders in different countries around the world. The data mapped here is sourced from United Nations Office on Drug and Crime (UNODC) and shows murder frequency in 207 countries around the world. The data and the map are explained here [https://www.theguardian.com/news/datablog/2011/oct/10/world-murder-rate-unodc].

Census 2010 Interactive Map: Texas Population by Race, Hispanic Origin To understand and appreciate this map you have to head over to the Texas Tribune website, https://www.texastribune.org/library/data/census-2010/

This interactive map lets users explore demographic trends in the state. By using the 'Select Map' feature at the top of this map, you can see the population density of different races in Texas. With the census coming up in Pakistan, this is an idea that can be easily applied back home. This map as well as the previous one have also been created through Fusion Tables, and if you follow the advanced tutorials given there you can make similar maps as well.



THINGS TO REMEMBER

1. You can also add photographs and images to your map that will appear when someone clicks a pointer. To do that, simply add an images column to your dataset and add pictures with each entry.
2. This tutorial touches the very basic of what you can do with Google Fusion tables. In real, the tool can be used to map much more extensively and to mesh different data sets together. Once you have got the hang of the basics, head over to Fusion Tables Gallery for tutorials and examples of advanced usage.



SECTION FIVE

ADDITIONAL RESOURCES

CHAPTER ELEVEN

AN ONGOING QUEST

TECHNOLOGY is what makes data and data journalism so relevant today. But technology is forever changing, evolving into new, better and faster forms. The data journalist thus, needs to stay on top of the related developments. In addition to the tools you are using for data journalism, it is also important to keep track of new trends in data journalism and stay tuned to the trendsetters in the field. In this chapter, we simply list out some additional resources that will help you add to your repository of tools that you are using as a data journalist. Happy learning to you all!

Tools

These are additional tools and applications that can help you work with your datasets.

1. PDF Join – allows you to merge several PDF files into one: <https://www.pdfjoin.com/>
2. Tabula – allows you to extract data from PDFs (doesn't work for scanned documents): <http://tabula.technology/>
3. CometDocs – allows you to extract data from PDFs: <http://www.cometdocs.com/>
4. ScraperWiki – allows you to extract data from PDFs and the web: <https://scraper-wiki.com/>
5. Datawrapper – <https://datawrapper.de/>
6. TileMill – a more advanced mapping tool than Google Fusion Tables – <https://www.mapbox.com/tilemill/>
7. Google Spreadsheets – do what you can with Excel, only online with Google. Access from your Google drive
8. Chrome plugin for scraping: <https://chrome.google.com/webstore/detail/scraper/mbigbapnycgaffohmbkdlecac-cepngjd?hl=en>

9. Automated data analysis: <https://www.statwing.com/>
10. Scraping for developers: <https://morph.io/>
All these tools have their own tutorials that you will find on their website.

Courses and Tutorials

1. Extracting tables from PDFs to Excel

The Electronic Data Resources Service has a detailed tutorial on extracting tables from PDFs to Excel. Download the tutorial from here – <http://www.library.mcgill.ca/edrs/Services/publications/howto/PDFtoXLS/PDFtoExcel.html>

2. Doing Journalism with data – First Steps, Skills and Tools

This is a great initial course for data journalists. You can revise some of the skills you have learned in this manual and learn a few new ones. You can also use practice material available through the course website and be introduced to some great examples of data journalism.

Tools

3. Bulletproof Data Journalism

Data is tricky, and numbers can be used to deceive. This course tells you about the traps you can fall into while working with data from different sources. It also discusses technical errors in data and shows you how to navigate complexities of working with data

This course is also free and self-paced and is accessible here – <http://bit.ly/2m15Rtj>

4. Managing Data Journalism Projects

This course will be especially helpful to those who work in more managing positions (think assignment desk). The course shows you how to manage a data journalism project, where to start and how to bring it to fruition. One thing that you will learn here, which we have not touched upon in this manual is creating effective teams for collaborative work on data based investigative projects.

Again, a free course, that you can take at your own pace here <http://bit.ly/2m13LE9>

5. School of Data

The School of Data aims to The school of data includes multiple customizable stand alone modules to develop various data journalism based skills. You can access it at <https://schoolofdata.org/courses/>

Books

1. The Data Journalism Handbook

The data journalism guidebook is an international, collaborative effort involving dozens of data journalism's leading advocates and best practitioners, including industry leaders like the BBC, Deutsche Welle, the Guardian, ProPublica, the Washington Post, and many others. The guidebook details various aspects of data journalism from an on ground perspective. A free online version of the guidebook can be accessed at <http://datajournalismhandbook.org/1.0/en/>

Communities

1. Data driven journalism – Facebook

An extension of <http://datadrivenjournalism.net>, the Facebook page <https://www.facebook.com/pg/data.driven.journalism> is a community of over 8,000 people, including data journalists from around the world. The page describes itself as a "part of an initiative to expand and strengthen an international network of data journalists, designers, developers, and others to encourage collaboration and to exchange expertise and best practices."

You can join the page to connect to other data journalists, share ideas and find inspiration.

Inspiration

1. Guardian Data Journalism Blog

From deep investigations to simple visual pleasures, subscribe to data journalism blog at The Guardian for a wealth of data based stories. In addition, the data blogs also keep all their data open, which means that you can download and use the data yourself to explore other stories.

Access Data Journalism Blog by The Guardian at <https://www.theguardian.com/media/data-journalism>

2. Bloomberg – Graphics

This site will offer new ideas and inspiration for the visual element of your data journalism. Vibrant, innovative and aesthetically interesting, Bloomberg graphics would definitely give you a few new ideas to work on.

Visit at <https://www.bloomberg.com/graphics>

3. Poynter Institute

Poynter brings together tools, tips tricks, trainings and research into the future of journalism. Data journalists interested in keeping up with the way future of journalism is shaping up should keep track of what Poynter has to offer at <http://www.poynter.org>

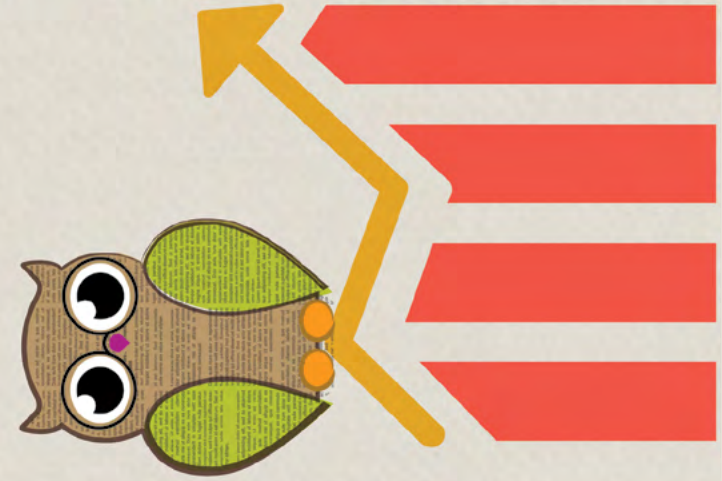
4. Driven by Data

This is a student project by City University London exploring what data journalism is, focus on particular case studies and produce our own visualisations from raw data. Visit at <https://drivenbydata.wordpress.com>

GLOSSARY

ADB	Asian Development Bank
CRTI	Coalition on Right to Information
ECP	Election Comission of Pakistan
FBR	Federal Bureau of Revenue
FIR	First Information Report
Fol	Freedom of Information
KP	Khyber Pakhtunkhwa
MNA	Member National Assembly
MPA	Member Provincial Assembly
PCP	Pakistan Code of Criminal Procedure
PILDAT	Pakistan Institute of Legislative Development and Transparency
PIPS	Pakistan Institue of Parliamentary Services
PSP	Pak Sarzameen Party
PTA	Pakistan Telecommunication Authority
RTI	Right to Information
SDG	Sustainable Development Goals
URL	Universal Resource Locator

ANNEXURES



ANNEX 1

Bank Challan form for submitting fees to National Bank for

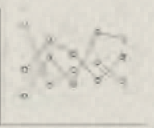
PROVINCIAL		Treasury of Sub-Treasury National Bank of Pakistan State Bank of Pakistan		Form No. 32-A Provincial/ Central	
Challan of cash paid in to Challan No.					
By Whom Tender	To be filled in the Remitters		Amount	To be Filled by the Department Officer or the Treasury	
	Name or designation and address of the person of who's behalf money is paid	Full particulars of the remittance and of authority (if any)		Head of account	Order to the Bank
		Total:-			
(in words) Rupees Received Payment Treasury/Officer Signature		Agent/Manager Date	To be used only in the case of remittances to bank through an officer of the Government		

PROVINCIAL		Treasury of Sub-Treasury National Bank of Pakistan State Bank of Pakistan		Form No. 32-A Provincial/ Central	
Challan of cash paid in to Challan No.					
By Whom Tender	To be filled in the Remitters		Amount	To be Filled by the Department Officer or the Treasury	
	Name or designation and address of the person of who's behalf money is paid	Full particulars of the remittance and of authority (if any)		Head of account	Order to the Bank
		Total:-			
(in words) Rupees Received Payment Treasury/Officer Signature		Agent/Manager Date	To be used only in the case of remittances to bank through an officer of the Government		

ANNEX 2

Charts do's & don'ts by Infogram

Line Chart



A line **chart** reveals trends or change over time. Line charts can be used to show relationships within a continuous data set, and can be applied to a wide variety of categories, including daily number of visitors to a site or variations in stock prices.

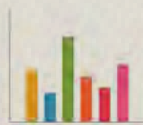
Do's:

- Start the x-axis at zero - Otherwise your viewer might draw inaccurate conclusion.
- Clearly label your axes - Make sure the viewer knows what they are evaluating.
- Remove distracting chart elements - Grids, varying colors, and bulky legends can distract the viewer from quickly seeing the overall trend.
- Zoom in on the y-axis if your data set starts above zero - In certain cases, changing the scale of the y-axis makes it easier for the viewer to see the trend.

Don'ts:

- Compare more than 5-7 lines - You don't want your chart to become cluttered or hard to read. Visualize the data you need to tell your story, nothing more.
- Use too many colors to distinguish the different data sets - Less is more. Try to use color to highlight one specific point.

Column and Bar Charts



Column and bar charts are used to compare different items, or show a comparison of items over time. Bars on a column chart are vertical while bars on a bar chart are horizontal. Bar charts are generally used to help avoid clutter when one data label is long or if you have more than 10 items to compare. Bar charts can also display negative numbers.

Do's:

- Start the y-axis at zero - Our eyes are sensitive to the area of bars on a chart. If those bars are truncated, the viewer might draw the wrong conclusions.
- Label the axes - Labeling the axes gives your viewer context.
- Put value labels on bars - This helps to preserve the clean lines of the bar lengths.

Don'ts:

- Use too many colors - Avoid the 'rainbow effect.' Using a single color, or varying shades of the same color, is a much better practice. You can highlight one bar in particular if that is the message you want to get across.
- Use 3-D imagery or distracting visuals - Adding unnecessary visuals can make your chart hard to read and a bit of an eyesore.

Pie Chart



Pie charts are used to show parts of a whole. A pie chart represents numbers in percentages, and the total sum of all the divided segments equals 100 percent.

Do's:

- Make sure your segments add up to 100 - Sounds obvious, but this is a common mistake.
- Use when comparing percentages people are familiar with - 25%, 75%, 100%.
- Keep it clean and consistent - Compare just a few categories to get your point across.

Don'ts:

- Compare too many categories - Nobody likes a messy pie chart.
- Use a pie chart if your segments end up being roughly the same size - This means your data is too hard to distinguish. Use a bar or column chart.
- Compare two pie charts side by side - A line or area chart would work better.
- Use 3-D imagery or tilt your pie chart - This often makes your data impossible to read, because your viewer is trying to quickly compare angles.

Dual Axis Chart



What happens when you want to compare two variables with different magnitudes and scales of measurement, in hopes of proving or disproving a trend? The answer is the dual axis (multiple axes) chart.

With a dual axis chart you are essentially combining multiple charts and adding a second y- axis for comparison.

Do's:

- Make sure your y-axes are related – Dual axis charts can be useful when comparing values that have different units of measure because the things they measure are somehow related.
- Place primary y- axis on the left – We are hard-wired to look at the y-axis on the left first.
- Use contrasting colors – Color code your dual axis chart to make it easier for people to understand the data sets you have plotted.

Don'ts:

- Compare similar units of measurement – Take this opportunity to show you've done your research and give diversity to your data.
- Include too many variables – Using multiple lines for your dual axis chart is considered acceptable, although it does run the risk of getting cluttered. We suggest you don't use more than four.

Scatter Plot



Scatter plots, a lot like line graphs, use horizontal and vertical axes to plot data points. But, scatter plots have a very unique purpose – they show how one variable affects another, meaning you can visualize distribution, relationships, and trends in your data.

The relationship between two variables is referred to as their correlation. The closer the data points come to making a straight line means the correlation between the two variables is higher, or the relationship is stronger. When reading a scatter plot you need to look at the direction, slope, and strength of the data points. The line that best fits the points on the scatter plot, or best fits the correlation, is known as the regression line.

Do's:

- Use scatter plots when you have a lot of data - Scatter plots lend themselves to large data sets.
- Use regression lines to guess Y if you have X - Helpful when estimating certain unknown data points.
- Start the Y-axis at 0 - If you set the Y- axis above 0 it skews the way your graph looks, and your data won't be presented accurately.

Don'ts:

- Compare too many regression lines - A scatter plot can get messy if you try to compare too many relationships/trends.
- Exclude outliers - Outliers may feel like they aren't adding to your story, but they are still an important part of the data set. They are useful when trying to fully understand the distribution of your data.

Area Chart



Area charts are a lot like line charts, with a few subtle differences. They can both show change over time, overall trends, and continuity across a dataset. But, while area charts may function the same way as line charts, the space between the line and axis is filled in, indicating volume.

Do's:

- Make it easy to read - Avoid occlusion. This happens when one or more layers covers important information on the chart.
- Use a stacked area chart - If you have multiple data sets and want to emphasize part-to-whole relationships.
- Use area charts to look at the bigger picture - Take population for example: Line charts are good for showing net change in population over time, while area charts are good for showing the total population over time.

Don'ts:

- Compare too many datasets - Use a line chart, its cleaner.
- Layer too many like colors - This makes your chart almost impossible to read.
- Make viewers do 'visual math' - Give the proper context with appropriate labels and legends.

121

Pyramid



Pyramid charts (triangle chart or triangle diagram) are a fun way to visualize foundation based relationships. They appear in the form of a triangle that has been divided into horizontal sections with categories labeled according to their hierarchy. They can be oriented up or down depending on the relationships they represent. The stacked layers can also show the order of steps in a particular process.

Do's:

- Pick a topic and clearly label your subcategories - Decide what information you want to convey with your pyramid and clearly label your layers.
- Organize your subcategories - Decide the order and value of each section on your pyramid.
- Organize the subcategories based on their hierarchy.
- Be consistent - Keep the spacing of your sections even and pick a pleasing color palette.

Don'ts:

- Include too many subcategories - Too many colors or layers can make your pyramid hard to read.
- Omit important subcategories - While you want to be selective and straightforward, you don't want to leave out important data just to suit the aesthetics of your pyramid.

Treemap



Treemaps show parts of a whole. They display hierarchical information as a cluster of rectangles varying in size and color, depending on their data value. The size of each rectangle represents a quantity, while the color can represent a number value or a category.

Treemaps allow you to view trends and make comparisons quickly – especially if one color is particularly prominent. While spreadsheets can show multiple rows of data, treemaps can accommodate hundreds of thousands of items in one organized display, making it easy to spot patterns in seconds. Plus, if made correctly, they make very efficient use of space.

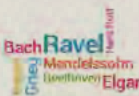
Do's:

- Start with clean data and a clear message - Treemaps can often involve a lot of data, so it's important to know exactly what you want to highlight.
- Use bright, contrasting colors so each region is easily defined - But, remember to avoid the 'rainbow effect.' Choose your colors wisely.
- Label each region appropriately with text or numbers - This makes it easier for the viewer to evaluate your treemap quickly, without error.

Don'ts:

- Label every box if you have a lot of data - Leave the smaller boxes empty if you feel you have too much text.
- Clutter your treemap with too many boxes - Treemaps can contain any number of boxes, but space is limited! You don't want your treemap to be hard to read.

Word Cloud



Word clouds (also known as tag clouds) are a type of weighted list. Word clouds display text in varying font sizes, weight, or colors to show frequencies or categories. They can be arranged alphabetically or at random. They help people identify trends and patterns that might have been difficult to see otherwise.

Do's:

- Provide context - Word clouds are visually eye-catching and provide information about frequency but they often don't give the viewer any context.
- Use word clouds to show frequency - Avoid using them to display complex topics like the budget or healthcare crisis.
- Watch your word length - Longer words take up more space and can be misleading.
- Use word clouds to discover new SEO - Word clouds are great for filtering and analyzing data.

Don'ts:

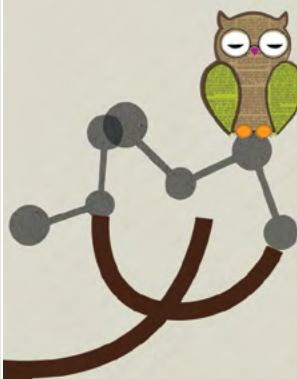
- Make your words too similar in size or color - Word clouds run the risk of being hard to analyze if the words all look the same.
- Use too many words - Word clouds can take up a lot of space
- Overuse the word cloud - Use wisely!

ABOUT THE AUTHORS

Sadaf Khan is a former broadcast journalist currently engaged in media rights advocacy and capacity building efforts in Pakistan. Her interests include media convergence, policy and internet governance. She has an MSc. in Media and Development from London School of Economics and is currently engaged with Media Matters for Democracy, Pakistan and serving as visiting faculty at Qaid-e-Azam International University Islamabad. She has been engaged with DW Akademie as a consultant for data journalism and Right to Information related initiatives.



Asad Baig is a techie by academics and a journalist by profession. He is currently working as a freedom of expression advocate, involved in various initiatives for technological development of media industry in Pakistan, journalist safety, recognition of freedom of expression as a media and a civil right and reforms of media regulatory systems in the country. He is the co founder of Media Matters for Democracy, Pakistan and is currently heading the organization as Executive Director. He has been engaged with DW Akademie as a consultant for data journalism and Right to Information related initiatives.



This manual was developed as part of the pilot project 'Implementing Right to Information Acts – Promoting Data-Driven Journalism'.

The project promotes dialogue between media professionals and the civil servants at public authorities. The aim is to implement the Right to Information Acts in Punjab and Khyber Pakhtunkhwa in a user-friendlier manner. Public Information Officers receive needs-based trainings on responding to requests for information. Technical staff at public authorities learns how to publish relevant data on digital platforms.

Journalists acquire the skills to make professional use of the provided data – on issues such as taxes, local planning and accountability – in their reporting. In workshops and by means of subsequent mentoring, media professionals learn how to carry out data-intensive research, data analysis and data visualization in order to prepare clear, data-based stories. The project also supports media house managers in developing workflows to integrate data-driven journalism into everyday editorial processes.

More information:
www.media-humanrights-governance.de



Implemented by

