**Data Journalism Test Question Bank**

These test questions accompany Internews’ 200-hour data journalism training program. The test is administered before and after each workshop. Answers are provided in **bold**. These sample questions come from the Data Journalism Certification program run in Pakistan in 2016 and 2017.

*Weeks 1 and 2*

1. What is the specific purpose of Google Advanced Search function “Site:” ?
	1. Allows you to search a specific date range
	2. Allows you to search for a specific filetype
	3. **Allows you to search within a specific web address**
	4. Allows you to search for content in a specific language
	5. Don’t Know
2. What is the specific purpose of using a hypothesis and questions for your investigation?
	1. **It ensures you focus on a coherent story instead of just a statistical overview**
	2. It allows you to follow the scientific method to produce accurate data findings.
	3. It allows you to prove what you already suspect to be true about an issue and ignore conflicting information.
	4. Don’t Know
3. Which of the following tools is NOT a basic function of Excel?
	1. Calculating sums
	2. Calculating averages
	3. **Scraping data**
	4. Sorting data
	5. Don’t Know
4. Which of the following is NOT a tool for scraping data?
	1. Zamzar.com
	2. **Google Fusion Tables**
	3. Firefox Table2Clipboard
	4. Google spreadsheets =ImportHTML
	5. Don’t know
5. Which describes why data journalism is important for the future of the media industry?

	1. It lets you tell faster breaking news
	2. It lets you tell shorter stories with less complexity.
	3. **It provides analytical, in-depth content**
	4. It is easier than traditional journalism
6. What information is NOT essential to include in the Source tab of your data file?

	1. **A clean copy of the data.**
7. The name of the organization that produced the data.
8. Any notes and explanations that accompany the data.
9. The date the data was produced and downloaded.
10. Which is NOT an important question to ask when you receive a data set?
	1. Who created this data?
	2. **Can I manipulate this data to support my political views?**
	3. What was the sample size for the data collection?
	4. Did the researchers used random sampling to choose participants?
11. Which correctly describes the data journalism process?

	1. Cleaning data, analyzing data, scraping data.
12. **Finding data, scraping data, cleaning data, analyzing data, visualizing data**.
13. Visualizing data, analyzing data, scraping data, cleaning data, combining data
14. Analyzing data, combining data, visualizing data.
15. Which is the strongest hypothesis for a data-driven story?

	1. The government does not provide enough education for mothers on the importance of vaccination so children are still being infected and dying of these diseases.
	2. The number of vaccine preventable diseases is growing in the poorest regions of Pakistan and the most effected group are boys under the age of five.
	3. **A lack of a comprehensive strategy that includes education, vaccine distribution and health staffing are contributing to the persistence of vaccine preventable diseases that have been eradicated in most of the world.**
	4. The government should do more to ensure all children receive basic vaccines.
16. Describe three things you hope your reader understands after reading a data-driven story about an issue:

…Causes of the problem, cost of the issue extent of the issue, demographics of those affected, cost of the issue, impact of the issue on citizens, a comparison of the issue across regions or demographics, potential solutions to the issue. ……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

1. Tell us which of the following spreadsheet skills you can now do on Excel.

\_\_\_ None.

\_\_\_ I can use spreadsheets to calculate totals

\_\_\_ I can use spreadsheets to calculate differences

\_\_\_ I can use spreadsheets to calculate percentage change

\_\_\_ I can use spreadsheets to calculate percentage difference

\_\_\_ I can use spreadsheets to calculate average / median (and what they are)

\_\_\_ I can use spreadsheets to calculate rate / ratio

\_\_\_ I can use spreadsheets for sorting & filtering

\_\_\_ I can use spreadsheets with pivot tables

\_\_\_ I can mport data into a spreadsheet

\_\_\_ Other:

1. List four sources of data for a data-driven story.

**Government, civil society, academia, international organizations, research organizations, survey companies, crowd-sourcing, first-hand data collection, sensors**.……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………… ……………………………………………………………………………………………..

*Week 3*

1. What is the best chart type to show the rate of maternal mortality in Pakistan, India and Bangladesh each year from 2005 to 2015?
	1. A bar chart
	2. **A line chart**
	3. A pie chart
	4. A map
	5. Don’t Know
2. What is an example of choosing bad proxy data?
	1. Not using random sampling to select participants in a poll about the reason their children are out of school.
	2. Reporting that one candidate is in the lead even if their respective polling scores are within the margin of error of each other.
	3. **Using the number of polio workers attacked to measure how effective a polio vaccination program has been.**
	4. Don’t Know
3. What will Excel return if you type in B4: =IF(B2="F","Female","Male").

In this case, B2 contains “Unknown.”

* 1. M
	2. Female
	3. **Male**
	4. Unknown
	5. Don’t Know
1. Which of the following is NOT a tool for visualizing data?
	1. Infogr.am
	2. CARTO
	3. DataWrapper
	4. **Open Refine**
	5. Don’t know
2. Which describes why data visualization is important for data journalism?

	1. It lets you replace articles with maps
	2. **It lets you reinforce complex comparisons**
	3. It lets readers come to their own conclusions
	4. It is easier than traditional journalism
3. Which is **NOT** a step before creating a pivot table?
	1. Ensure column headers start in Row A.
	2. **Add filters to the data table**
	3. Split columns that blend two different data types.
	4. Create a copy of the original data set.
4. Which is the best headline for a data visualization?

	1. Under five child death rates fall.
5. GPD growth in 2015 in South Asia
6. UNESCO child dropout rates in South Asia
7. **Pakistan leads region in economic growth**

1. Choose out of the four correlation examples below, which is an example of causation?

	1. Girls drop out of secondary school in regions where there are no bathroom facilities in the schools.
	2. Mosquito-borne disease incidences are growing in regions of Pakistan where rainfall is rising.
	3. **Several scientific studies prove chemical pesticides cause birth defects among farmers not using proper protection.**
	4. Crime is rising in neighborhoods with large immigrant populations.
2. List three things reasons to include a data visualization in your story: **(3)**
* **Makes the data easier to understand and compare**
* **Grabs the audience’s attention**
* **Allows the audience to explore the data**
* **Tells a story about the data**
1. List four design errors in this graph: **(4)**

* **Headline doesn’t tell a story**
* **Too many colors and fonts**
* **3D effect distorts data**
* **Direct labeling is not used**
* **Too many pie slices so a bar chart should be used**
* **Slices are too close in size that you can’t make a visual comparison**
* **No source is listed**
1. List four different functions in Excel (with the formula name). **(4)**
* **Sum =SUM**
* **Average =AVERAGE**
* **Median =MEDIAN**
* **Join tables =VLOOKUP**
* **Trim spaces =TRIM**
* **Division =A2/B2**
* **Multiplication = =A2\*B2**
* **Percentage =part/whole**
* **Percentage change =(new number-old number)/old number**
1. Under what circumstances would you use the following formula? =TRIM(B2)

**To remove extra blank spaces**

*Week 4*

1. What is the best visualization type to compare how much each state spends on maternal healthcare compared to its maternal death rate when there are no geographical trends?
	1. A line chart
	2. A bubble map
	3. Two choropleth maps
	4. **A grouped bar chart**
	5. Don’t Know
2. What is the best example of a data-driven question for a data-driven accountability interview?
	1. You lost your baby during childbirth like almost 1 out of 5 women in this province. Can you tell us what the barriers were to getting professional help for your birth?
	2. You have been studying trends in maternal healthcare in this province for decades, can you tell us why most mothers seek medical care for the birth of the first child but according to our data, very few seek ante-natal care?
	3. **You are responsible for the budgeting for maternal healthcare in this district. Can you tell us why there are only three incubators even though pre-mature birth is the top cause of death in this district?**
	4. Don’t Know
3. What is the best example of a number lead?
	1. 1,200 are arrested every month for speeding.
	2. **Three in four women who are sexually assaulted never report it.**
	3. 15 more female candidates needed for Parliament.
	4. One out of two babies are born prematurely.
	5. Don’t Know
4. Which **NOT** important when writing your story?
5. Writing with a sense of the audience as citizens who, armed with information, can take action.
6. Writing in a conversational voice, including the use of an empathetic character.
7. Writing at a slow enough pace to promote comprehensibility and learning.
8. **Use of technical terminology to ensure your story sounds serious.**
9. Which does NOT describe a goal of data stories?

	1. Generate widespread public debate
	2. Overcoming indifference or inertia by politicians and donors.
	3. **Present complex problems so others can explore solutions.**
	4. Increase demand for more data from government by citizens for better governance
10. Which is NOT a step before creating a Tableau Visualization?
	1. Ensure column headers start in Row A.
	2. **Add filters to the data table**
	3. Split columns that blend two different data types.
	4. Create a copy of the original data set.
11. Which is the best headline for a data visualization?

	1. Measles rates in South Asia over 10 years.
12. Increased health spending in Pakistan.
13. **One out of four girls stays out of school to care for parents.**
14. Pakistan lags behind in vaccination rates.

1. When is the correct time to create a visualization or map for your story?
	1. Right after doing some preliminary research
	2. **Right after writing the story**
	3. Before creating a sketch and storyboard for the visualization
	4. When defining the focus of the story
2. Write a simplification of this data:

*Pakistan has a literacy rate of 58 percent, which has improved from 35 percent in 1990-91, but still way behind the Millennium Development Goals (MDG) targeted at 88 percent, which was to be achieved by the end of 2015.*

……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

1. List four design and storytelling errors in this graph:



……………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

1. List the four data-driven interview types.

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

1. Under what circumstances would you use the following formula? =(C2-B2)/B2

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

*Week 5*

1. What is the best example of a data-driven hypothesis?
	1. If Pakistan invested more in infrastructure, agriculture would be more profitable.
	2. Pakistan’s GDP would grow if more businesses paid taxes the way they do in Western Europe.
	3. **Pakistan cannot increase agricultural exports of wheat until it makes enriched seeds and fertilizer as well as low cost irrigation subsidies available to farmers.**
	4. Pakistan could improve living standards for citizens in remote areas by ensuring telecom access to impoverished citizens and reducing levels of poverty in the rural parts of the country.
2. What is the best lead for a story about contraceptive use in Pakistan?
	1. The top three reason most women in Pakistan don’t use contraceptives are family pressure, fear of side effects and lack of access.
	2. **Young women in Pakistan are so afraid of displeasing their families that they endanger the wellbeing of the children they already have to avoid contraceptives.**
	3. Contraceptive rates in Pakistan are lower than in other South Asian countries, leading to faster population growth and stretched resources.
	4. Pakistan’s contraceptive rate will not increase until the government removes the primary barriers to access for women in rural areas.
3. Rank the simplification of numbers from best to worst (1 is the worst, 4 is the best)?

\_\_3\_Countries with more educated girls, such as Sri Lanka and India have more employed women while countries with low rates of girls' education, such as Pakistan and Afghanistan, have low rates of women's employment.

\_\_1\_Women in Sri Lanka and India have high rates of girls' education while women in Pakistan and Afghanistan have low rates of girls' education. Women in Sri Lanka and India have medium rates of employment while women in Pakistan and Afghanistan have low rates of employment.

\_\_2\_Women in Afghanistan and Pakistan have low rates of girls' education and low rates of employment for women. Girls in Sri Lanka and India have high rates of education and high rates of employment among women.

\_\_4\_Girls' access to healthcare and basic education corresponds to higher levels of employment and financial stability among women in South Asia.

1. Whichis the **LEAST** important use of data in your data-driven solution story?
	1. **Measuring the revenue lost by the problem.**
	2. Measuring the impact of the problem on specific groups.
	3. Measure the causes of the problem.
	4. Measuring the effectiveness of proposed solutions.
2. Which is a **good** example of a solution interview question?

	1. How can we solve the problem of low contraceptive use among Pakistani women?
	2. Why do two in three women not use contraceptives in Pakistan?
	3. Which countries have done a better job than Pakistan in increasing contraceptive use among women?
	4. **What programs have had the most success in boosting contraceptive rates in the country and how did you measure success?**
3. Label from 1 to 4 the correct order of steps in creating a Tableau Visualization?

\_4\_Create the dashboard in Tableau.

\_1\_Sketch the visualization on paper.

\_2\_Format the data for upload to Tableau.

\_3\_Create a worksheet in Tableau.

1. Which is NOT an objective of a lead for a data story?
	1. Show the impact of the numbers.
	2. Prioritize the issue over the number.
	3. Surprise the audience .
	4. **Introduce the topic without findings.**
2. What is the common mistake made in this number lead?

*Amal is among three in four Pakistani girls who never had a chance to go to primary school and now she is among four in five women who is unemployed and is among the 42% poorest people in the country.*

* 1. The numbers are not simplified so people have no way of knowing whether the numbers are good or bad.
	2. **There are too many numbers**
	3. There is no human angle illustrating the human impact of the data.
	4. There are too many technical terms so the meaning of the numbers is not clear.
1. Write the most important findings in more accessible language (maximum 3 sentences):

*Although Pakistan was one of the first countries in Asia to launch national family planning programs, current modern contraceptive use stands at only 26% with a method mix skewed toward short-acting and permanent methods. Most of the women in this study had low socio-economic status and were younger than 30 years of age. Four-fifths of the women consulted private sector health facilities for reproductive health services; proximity, availability of services, and good reputation of the provider were the main predicators for choosing the facilities. Husbands were reported as the key decision maker regarding health-seeking and family planning uptake. Overall, the current contraceptive use ranged from 17% to 21% across the districts: condoms and female sterilization were widely used methods. Woman’s age, husband’s education, wealth quintiles, spousal communication, location of last delivery, and favorable attitude toward contraception have an association with current contraceptive use.*

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

1. For each story structure, provide a topic and explain why that structure would be appropriate:

**Chronological:**

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

**Compare and contrast:** ………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

**Step-by-step through a problem or issue:**

………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………