

# Corpus Summary and Evaluation Materials

Alan Lundgard and Arvind Satyanarayan

Supplemental Material for *Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content*. IEEE Transactions on Visualization and Computer Graphics (TVCG), Special Issue on the 2021 Visualization Conference (VIS).



## CONTENTS

<b>1</b>	<b>Corpus Summary</b>	<b>2</b>
1.1	Descriptive Statistics . . . . .	2
1.2	Visual Fingerprint . . . . .	2
<b>2</b>	<b>Evaluation Design</b>	<b>3</b>
2.1	Rank-Choice Interfaces . . . . .	3
2.2	Rank-Choice Questions . . . . .	3
<b>3</b>	<b>Evaluation Questionnaire</b>	<b>11</b>
3.1	Demographic Questions . . . . .	11
3.2	Visualization Questions . . . . .	12

---

• Alan Lundgard is with MIT CSAIL. E-mail: [lundgard@mit.edu](mailto:lundgard@mit.edu).  
• Arvind Satyanarayan is with MIT CSAIL. E-mail: [arvindsatya@mit.edu](mailto:arvindsatya@mit.edu).

# 1 CORPUS SUMMARY

See Paper Section 3.2 Gathering A Corpus.

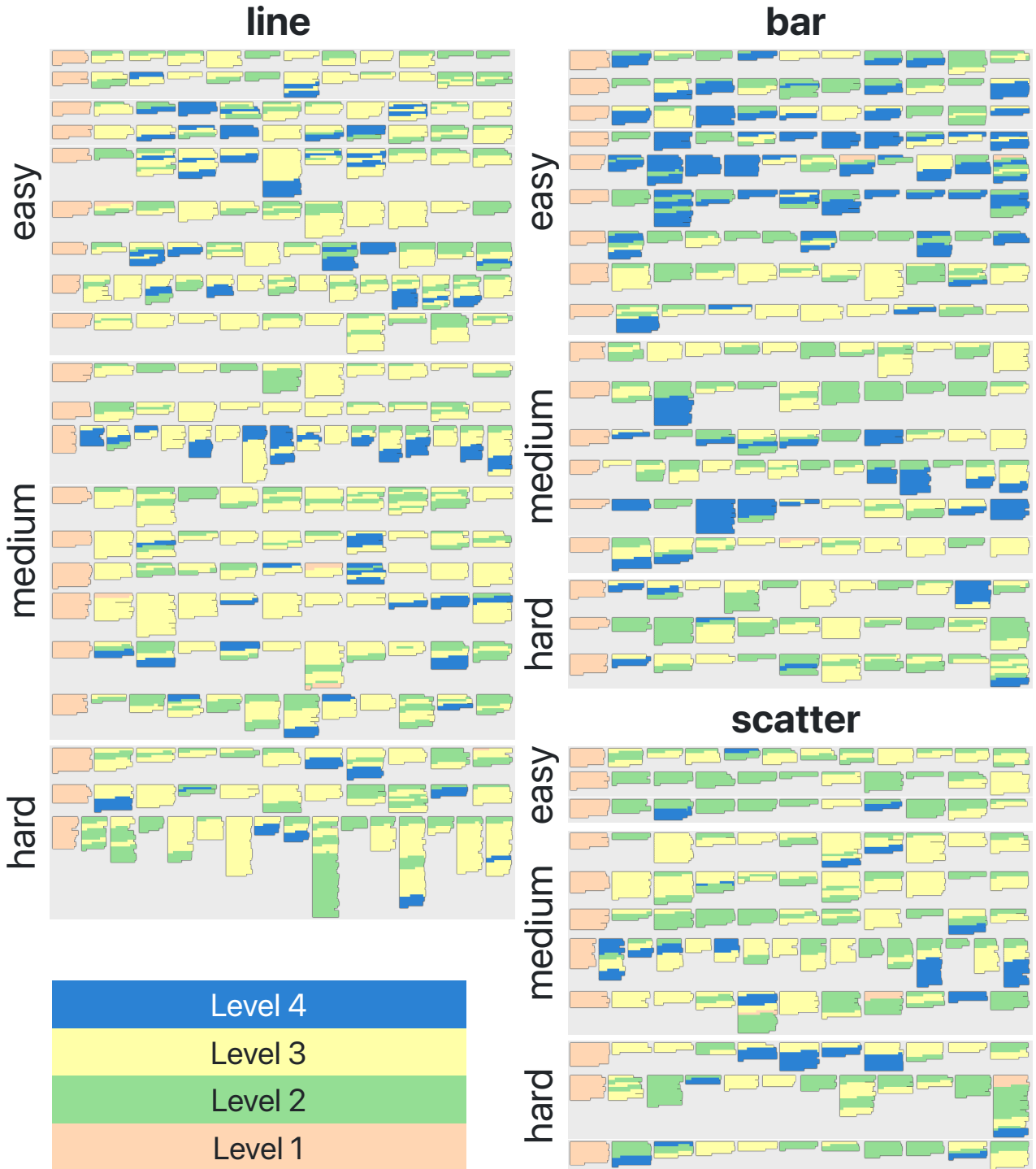
## 1.1 Descriptive Statistics

Our corpus consists of 582 visualization descriptions, authored by 120 participants. On average, participants wrote 3.7 sentences per visualization, resulting in 2,147 sentences total. The distribution of sentences across the four levels of semantic content is as follows.

Level 1: 9.1%, Level 2: 35.1%, Level 3: 42.9%, Level 4: 12.9%.

## 1.2 Visual Fingerprint

A visual “fingerprint” of our corpus, faceted by chart type and difficulty. Each row corresponds to a single chart. Each column shows a participant-authored description for that chart, color coded according to our model. The first column shows the provided Level 1 prompt.



## 2 EVALUATION DESIGN

See Paper Section 5.1 Evaluation Design.

### 2.1 Rank-Choice Interfaces

#### Interface for Blind Readers

Ranking Task (#3330)  
Suppose that you are reading a newspaper article about the number of COVID-19 tests and cases in different countries and continents of the world. You encounter a visualization that is inaccessible via screen reader. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

Here are the five textual descriptions for Ranking Task (#3330). Please rank each of the five textual descriptions below from most useful (1) to least useful (5).

1 2 3 4 5

Poland, Romania and Lithuania have the largest positive change in GDP per capita. Greece has the largest negative change in GDP per capita. Countries that already have a high GDP per capita have lower change in GDP per capita.  1  2  3  4  5

In most continents, there is a strong positive correlation between the total number of tests conducted, and the number of COVID-19 cases: the higher the total of tests, the higher the number of cases, and vice-versa. The countries with the lowest number of tests and cases appear to be in Africa. In general, the countries with the highest number of tests and cases appear to be in North America and Europe.  1  2  3  4  5

The number of COVID-19 tests ranges from 10,000 to 1 million. The number of confirmed cases ranges from 61 to 100,000. The USA has the highest number of tests and the highest number confirmed cases. Bolivia has the lowest number of tests and about 1000 confirmed cases. Uganda has a low number of tests and the lowest number of cases, about 100.  1  2  3  4  5

This is a scatter plot entitled "Total COVID-19 tests conducted vs. Confirmed cases" that plots the total tests by the total confirmed cases of COVID-19. The total tests are plotted on the vertical y-axis from 10,000 to 1 million. The total confirmed cases are plotted on the horizontal x-axis from 61 to 100,000 using a logarithmic scale. The plotted countries are color coded by their continent: Africa (gold), Asia (green), Europe (blue), North America (red), Oceania (purple), South America (burgundy).  1  2  3  4  5

Overall, countries that are considered to be wealthier (such as the United States) have been able to conduct more tests and report more confirmed COVID-19 cases. Countries in Africa have done the fewest COVID-19 tests, but this can't confirm that they also have the lowest number of cases, which may be unknown due to their relatively limited testing.  1  2  3  4  5

The rank-choice interfaces for blind and sighted readers. Blind readers were not shown the visualization, but instead shown a question prompt indicating the hypothetical context of the inaccessible chart. Sighted readers were shown the visualization, as well as the hypothetical context of the chart. Readers assigned usefulness rankings to each of the 4 descriptions (presented randomly) by selecting corresponding radio buttons, labeled 1 (least useful) to 4 (most useful). Additionally, we included a 5th choice as an "attention check" (i.e., a description whose content was irrelevant to the chart) to ensure that participants were reading each choice prior to ranking them.

#### Interface for Sighted Readers

Ranking Task (#3330)  
Suppose that you are reading a newspaper article about the number of COVID-19 tests and cases in different countries and continents of the world. You encounter the following visualization.

**Total COVID-19 tests conducted vs. Confirmed cases**

Here are the five textual descriptions for Ranking Task (#3330).

Which descriptive content below would be most useful to include in a textual description of the visualization? Please rank each of the five textual descriptions below from most useful (1) to least useful (5) by selecting the corresponding radio button that appears after each description.

1 2 3 4 5

Overall, countries that are considered to be wealthier (such as the United States) have been able to conduct more tests and report more confirmed COVID-19 cases. Countries in Africa have done the fewest COVID-19 tests, but this can't confirm that they also have the lowest number of cases, which may be unknown due to their relatively limited testing.  1  2  3  4  5

In most continents, there is a strong positive correlation between the total number of tests conducted, and the number of COVID-19 cases: the higher the total of tests, the higher the number of cases, and vice-versa. The countries with the lowest number of tests and cases appear to be in Africa. In general, the countries with the highest number of tests and cases appear to be in North America and Europe.  1  2  3  4  5

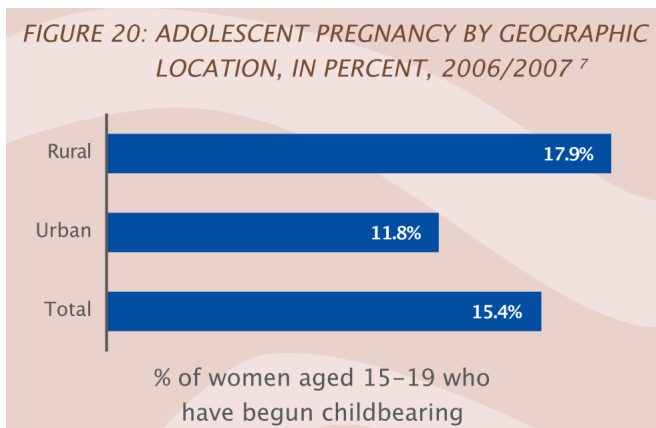
Poland, Romania and Lithuania have the largest positive change in GDP per capita. Greece has the largest negative change in GDP per capita. Countries that already have a high GDP per capita have lower change in GDP per capita.  1  2  3  4  5

The number of COVID-19 tests ranges from 10,000 to 1 million. The number of confirmed cases ranges from 61 to 100,000. The USA has the highest number of tests and the highest number confirmed cases. Bolivia has the lowest number of tests and about 1000 confirmed cases. Uganda has a low number of tests and the lowest number of cases, about 100.  1  2  3  4  5

This is a scatter plot entitled "Total COVID-19 tests conducted vs. Confirmed cases" that plots the total tests by the total confirmed cases of COVID-19. The total tests are plotted on the vertical y-axis from 10,000 to 1 million. The total confirmed cases are plotted on the horizontal x-axis from 61 to 100,000 using a logarithmic scale. The plotted countries are color coded by their continent: Africa (gold), Asia (green), Europe (blue), North America (red), Oceania (purple), South America (burgundy).  1  2  3  4  5

### 2.2 Rank-Choice Questions

1111\_bar\_easy\_academic.c.png



#### Question Prompt for Blind Readers

Suppose that you are reading an academic article about adolescent pregnancy by geographic location. You encounter a visualization that is inaccessible. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

#### Ranking Choices

**Level 1** This is a horizontal bar chart entitled "Adolescent pregnancy by geographic location, in percent, 2006/2007". Geographic location is plotted on the vertical y-axis for the locations Rural, Urban, and Total. The percent of women aged 15-19 who have begun childbearing is plotted on the horizontal x-axis for Rural, Urban, and Total.

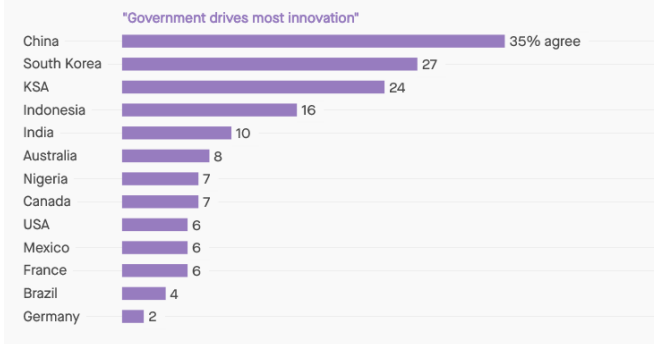
**Level 2** The percentage of pregnant adolescents is higher in Rural locations than in Urban ones. Rural women have the highest rate of pregnancy at 17.9%. Urban women have the lowest rate of pregnancy at 11.8%. The Total rate of pregnancy is 15.4%.

**Level 3** The percentage of adolescents who have begun childbearing in Rural locations corresponds to almost one fifth. The percentage of adolescents who have begun childbearing in Urban locations corresponds to almost one eighth. The percentage of women childbearing in Rural areas is higher than average. The percentage of women childbearing in Urban areas is lower than average.

**Level 4** There are more cases of adolescent pregnancy in Rural areas possibly because there is less attention to unwanted pregnancy in those areas, and because people might be less informed about how to avoid pregnancy. This could be a reflection of the lack of education or entertainment available in Rural areas. The 17.9% Rural pregnancy rate of 15-19 year olds is shockingly high.

**Attention Check** The house value of San Francisco has gradually increased year by year in a quasi-linear manner. A minimum of 20% of houses costs more than 1 million in the year 2012 and after 4 years the house costs of more than 70% in the year 2016.

## China leads in government support for innovation



### Question Prompt for Blind Readers

Suppose that you are reading an article in a business magazine about how people in different countries perceive their government's influence on technology innovation. In the course of reading this article, you encounter a visualization that is inaccessible. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

### Ranking Choices

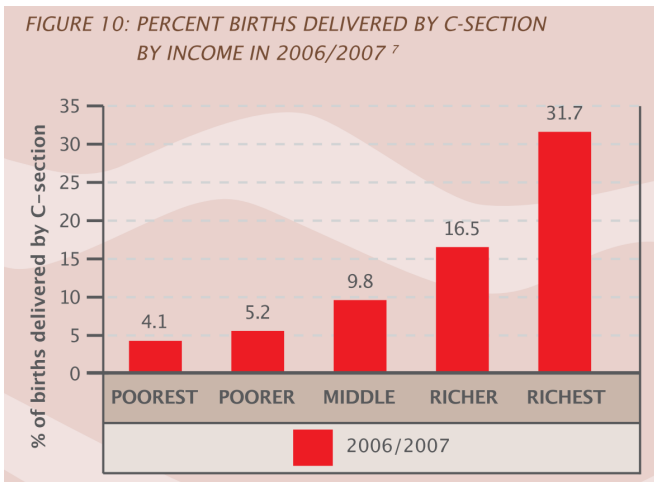
Level 1 This is a horizontal bar chart entitled "China leads in government support for innovation" that plots by Country the percentage agreement with the following statement: "Government drives most innovation". Country is plotted on the vertical y-axis. The percentage agreement is plotted on the horizontal x-axis.

Level 2 China has the highest agreement (35%) while Germany has the lowest (2%). At 6%, the USA has less agreement than China. Agreement in South Korea (27%) is second only to China.

Level 3 China is the country that has the highest agreement with the statement "Government drives most innovation." But this is still pretty low: only 35% of people in China agree with the statement. And there is a very sharp drop-off in agreement for the statement between China (the highest agreement) and Germany (the lowest agreement at 2%).

Level 4 In Asian countries, more people agree that their government drives innovation, although it is still a relatively small percentage for some (between 10 and 35% agreement). Countries in areas of the world other than Asia more strongly disagree that their government drives innovation (between 2 and 8% agreement). European countries seem to have the least agreement with the statement that their government drives innovation. Germans disagree with this statement the most (only 2% of Germans agree).

Attention Check Countries in Europe have a faster Netflix download speed. Countries in South America have a slower Netflix download speed. Countries in North America and Oceania have an average Netflix download speed.



### Question Prompt for Blind Readers

Suppose that you are reading a medical journal article about the number of births delivered by C-section depending on income levels. You encounter an inaccessible visualization. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

### Ranking Choices

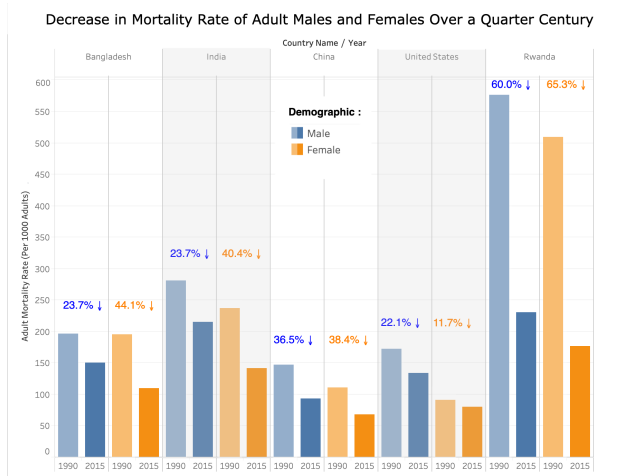
Level 1 This is a vertical bar chart entitled "Percent births delivered by C-Section by income in 2006/2007". The percent of births delivered by C-section is plotted on the vertical y-axis from 0 to 25. The income categories are plotted on the horizontal x-axis for the following categories: Poorest, Poorer, Middle, Richer, and Richest.

Level 2 The Richest have the highest percent of births by C-section (31.7%). The Poorest have the lowest percent of births by C-section (4.1%). The Richest are almost twice as likely to have a baby delivered by C-section than the Richer and 3x more likely than the Middle.

Level 3 Overall, the richer you are the more likely you are to have a baby delivered by C-section in 2006/2007. There is an exponential rise in babies delivered by C-section as you move up the income ladder in 2006/2007. There is a discrepancy of around 27% between the Poorest and the Richest people, and a nearly-doubled increased percentage of births delivered by C-section between the Richer and the Richest people.

Level 4 The clear trend for the Richest people to have births delivered by C-section suggests that income is a huge factor in determining whether births are delivered by C-section or not. This may link to the possibility for the Richest people to access and afford a better and top quality health system. The Poorest people have the lowest percentage of C-sections, likely because they couldn't afford it, or because they were treated differently. This suggests there is a poverty gap in access to and education in birthing choices and healthcare.

Attention Check 2016 has been the year with the highest number of sales for Walmart. Before 2016, 2015 seems to have had the biggest tendency for growth. 2011 is the year with least gains, and even so shows a decrease in the earnings.



**Question Prompt for Blind Readers**

Suppose that you are reading an academic article about the mortality rate by sex in different countries. You encounter a visualization that is inaccessible. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

**Ranking Choices**

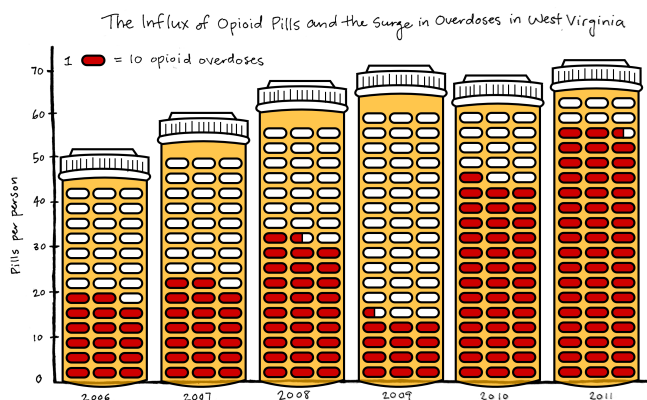
Level 1 This is a vertical grouped bar chart entitled "Decrease in Mortality Rate of Adult Males and Females Over a Quarter Century". The y-axis plots the Adult Mortality Rate (per 1000 Adults) from 0 to 600 with an increment of 50. The x-axis shows repeating bar-pairs for the years 1990 and 2015 by country and demographic (male or female).

Level 2 In India and Bangladesh, female mortality has decreased by a higher percentage (around 40%) when male mortality has only decreased by almost half that percentage (23%). On the other hand, in the United States female mortality has only decreased by 11% while male mortality has decreased by double that amount at 22%. China shows a significant decrease in mortality rate of 36% for males and 38% for females. The lowest mortality rate in 2015 for both males and females was in China.

Level 3 Every country's adult mortality rates for both males and for females have decreased over the course of 25 years from 1990 to 2015. For all countries except the USA, this is more marked among females than males. In the USA, this is the reverse, and the mortality rate among males has dropped at double the rate than among females. Of the countries shown, Rwanda's death rate was particularly high for both males and females, but has dropped considerably for both, putting its current mortality rate on a par with the other countries.

Level 4 The mortality rates decrease and the percentages are lower in countries such as the USA and China where they have fairly advanced healthcare. Bangladesh and India mortality rates are roughly the same even though their populations are vastly different. The percent decrease in male mortality rate is exactly the same in Bangladesh and India, even though India has far more people in their country. Overall, we can see that from 1990 to 2015 every country exhibits a decrease in the adult mortality rate. This decrease in mortality rate of both sexes might be explained by an increase in standard of living and development of medicine and technology.

Attention Check Belgium, Luxembourg, and the Netherlands have the most Netflix download speeds in peak hours, averaging about 4 Mbps, while Jamaica, Ecuador, and Costa Rica have the lowest Netflix download speeds at peak hours, averaging about 2 Mbps.



**Question Prompt for Blind Readers**

Suppose that you are reading a newspaper article online that is about the influx of opioid pills and the surge in overdoses in West Virginia. You encounter an inaccessible visualization. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

**Ranking Choices**

Level 1 This is an illustrated bar chart entitled "The Influx of Opioid Pills and the Surge in Overdoses in West Virginia" that plots the number of pills per person and the number of overdoses by Year. The Number of pills per person is plotted on the vertical y-axis from 0 to 90. The number of opioid overdoses is represented by each red pill equaling 10 opioid overdoses. The Year is plotted on the horizontal x-axis from 2006 to 2011.

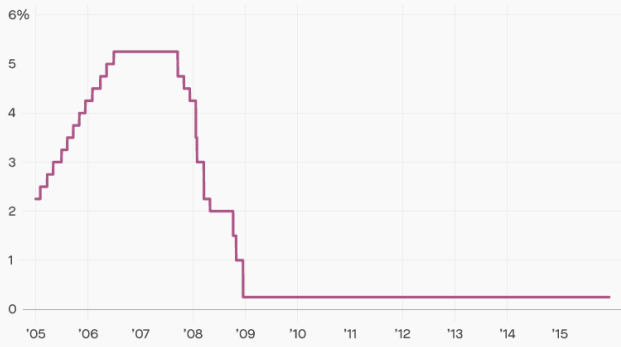
Level 2 In West Virginia, there is a positive correlation between the number of opioid pills per person and the number of overdoses per year. 2009 was an exception to this trend. 2006 was the year with the lowest number of pills per person. 2009 and 2011 were the years with the highest number of pills per person. Although 2009 was a year with one of the highest number of pills per person, the number of opioid overdoses was the lowest. 2011 was the year with the highest number of opioid overdoses.

Level 3 On average, there has been a steady increase in both the number of opioid pills per person and the number of opioid overdoses in West Virginia. While there is a steadily increasing trend in opioid overdoses per year, there was a dip in 2009, which indicates fewer overdoses despite the number of opioid pills per person continuing to increase.

Level 4 Apart from the number of opioid pills per person, there are probably many other socioeconomic factors that need to be considered when looking into the number of opioid overdoses in West Virginia. For example, the number of opioid overdoses could be related to the fact that each year, people are becoming more stressed due to other societal factors, such as unemployment.

Attention Check The highest number of people to be abused by the South African police force was between 2010-2011 with around 800 victims. The year with the fewest number of people was 2002-2003 with an estimated 330 victims.

### Federal Reserve federal funds funds interest rate target



#### Question Prompt for Blind Readers

Suppose that you are reading a business magazine article about the United States Federal Reserve interest rate target over the years. You encounter an inaccessible visualization. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

#### Ranking Choices

Level 1 This is a line plot entitled "Federal Reserve federal funds funds [sic] interest rate target" that plots the interest rate target by Year. The interest rate target is plotted on the vertical y-axis from 0 to 6%. The Year is plotted on the horizontal x-axis from 2005 to 2015.

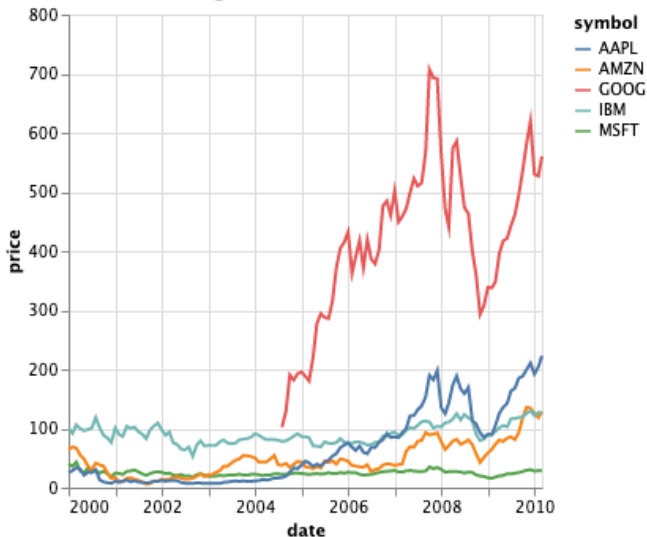
Level 2 From 2005 to 2007 the interest rate target increased from approximately 2 to 5 percent. From 2007 to 2008 it stayed the same at around 5 percent. From 2009 to 2015, the interest rate target remained flat at around 0 percent.

Level 3 The Federal Reserve interest rate rapidly increased from 2005 to 2007, going from 2 percent to 6 percent, reaching a plateau at 6 percent between 2006 and the end of 2007, and then sharply decreasing between 2008 and 2009, falling to almost 0 percent. From 2009 to 2015, the interest rate stays steady at around 0 percent.

Level 4 Before the 2008 financial crisis, the Federal Reserve interest rate was constantly increasing. Then, because of the financial crisis, the interest rate plummeted to almost 0 percent from 2009 onwards. This is because the Federal Reserve had to provide liquidity to the market.

Attention Check Between 2006 and mid 2008, Taiwan's exports in US dollars ranged between an estimated 22% and 5%. After that, there was a big drop in 2009 where the exports dropped to -38%.

### Big Tech Stock Prices



#### Question Prompt for Blind Readers

Suppose that you are reading a business magazine article about how the stock prices of Big Tech corporations have changed over the years. You come across a visualization that is inaccessible. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

#### Ranking Choices

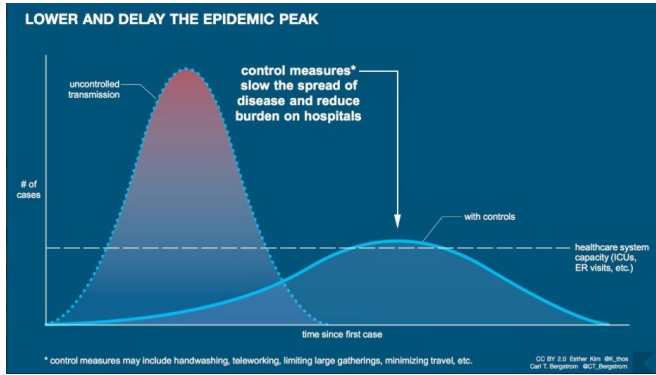
Level 1 This is a multi-line chart entitled "Big Tech Stock Prices" that plots the stock price of big tech corporations by year. The corporations include AAPL (Apple), AMZN (Amazon), GOOG (Google), IBM (IBM), and MSFT (Microsoft). The years are plotted on the horizontal x-axis from 2000 to 2010 with an increment of 2 years. The prices are plotted on the vertical y-axis from 0 to 800 with an increment of 200.

Level 2 Since mid-2004, GOOG had a steep increase in its stock price, and has the highest stock price since then. Beneath GOOG, AAPL has the second highest stock price since 2007, followed by AMZN, and then MSFT. MSFT has had the lowest stock price since around mid 2004, and there has been little change in its stock price over the years.

Level 3 In mid-2004, GOOG shares surged up, permanently exceeding the stock prices of the other four companies. GOOG took off from 2004 to 2008 and then took a big hit in 2008. All the other stock prices stay mostly steady for ten years, gradually rising, but nowhere close to as much as GOOG's dramatic rise. IBM has almost had the same value over the years, with a slight depression from 2002 to 2007 but still keeping its value around 100. AAPL started very low, started growing in 2004, and after that it kept a pattern of fluctuating growth. AAPL's highest stock price was 215 in 2010, and its lowest was around 10 in 2003. AMZN started a bit bigger but went low over the course of 2000. Then, it started growing in 2003 and has since followed a fluctuating growth pattern like AAPL. AMZN's highest stock price was 120 at the end of 2019, and its lowest was around 0 in the middle of 2001. MSFT has had no particular growth or drop and has kept its price at around 10 from 2000 to 2010.

Level 4 From 2000 to 2004 the Google stock price did not exist at all, due to the fact that the company didn't exist yet. Then, Google stock prices surged somewhere around 2005, possibly due to the success of YouTube. It was also around that time when the Google search engine became dominant. All companies' stock prices dipped significantly in 2007 to 2008 due to the global financial crisis, but bounced back shortly afterwards. Apple stock prices started to rise after 2009, possibly related to the iPhone's growth in popularity. In general, the corporations that manufacture computer hardware have more stable stock prices over the years, while the corporations that produce new Internet applications have more unstable and fluctuating stock prices.

Attention Check The highest US unemployment rate was registered in 1983, with a percentage of almost 11%. Before 1983, the highest US unemployment rate registered was 9% in the 1970s. After 1983, the US unemployment rate started lowering until it reached a percentage of around 5% in 1990.



#### Question Prompt for Blind Readers

Suppose that you are reading a newspaper article on the COVID-19 pandemic and you encounter the following “flatten the curve” visualization. However, the visualization is inaccessible. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

#### Ranking Choices

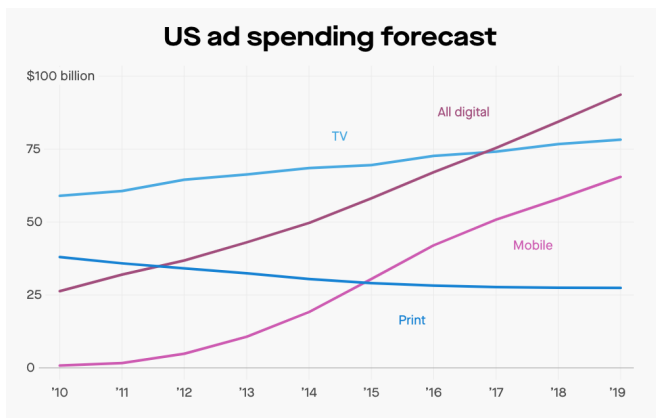
Level 1 This is a multi-line chart entitled “Lower and Delay the Epidemic Peak”. The chart plots the number of cases by the Time Since First Case. The number of cases is plotted on the vertical y-axis. The Time Since First Case is plotted on the horizontal x-axis. One of the lines is dotted and represents “uncontrolled transmission”. Another line is solid and represents “with controls”. A horizontal dashed line represents “healthcare system capacity (ICUs, ER visits, etc.)”. An annotation in the chart reads “control measures slow the spread of disease and reduce burden on hospitals.”

Level 2 The chart shows two possible extremes of the rate of rise and decline of COVID-19 cases. The two possible extremes (controlled and uncontrolled) are presented as two different lines.

Level 3 If the COVID-19 transmission is uncontrolled, the epidemic ends sooner. If the COVID-19 transmission is controlled, the epidemic ends later. If the transmission is uncontrolled, there are more simultaneous cases. If the transmission is controlled, there are fewer simultaneous cases. If the transmission is uncontrolled, the number of cases exceeds the capacity of the healthcare system. If the transmission is controlled, the healthcare system can support all the cases. In other words, the use of control measures allows the healthcare system to support all of the people who contract COVID-19 over time, although the epidemic may last longer as a result.

Level 4 The purpose of the chart is not to provide exact numbers, but to communicate to the public that there are two main ways the current crisis can play out: with and without control measures. The chart conveys that control measures provide security and protection to all members of society by not exceeding the capacity of the healthcare system (ICUs, ER visits, etc.).

Attention Check In June 2019 the auto industry employment hit the rock bottom. Since June 2019 employment in the auto industry had a steep rise.



#### Question Prompt for Blind Readers

Suppose that you are reading a newspaper article about advertisement spending in the United States. You encounter a visualization that is inaccessible. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

#### Ranking Choices

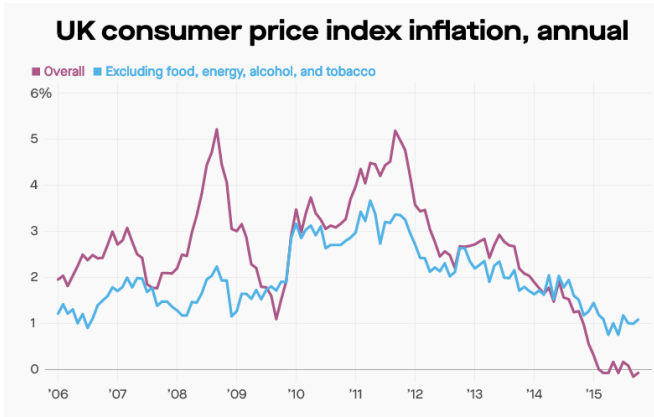
Level 1 This is a multi-line chart entitled “US ad spending forecast”. The chart plots ad spending by year. Ad spending is plotted on the vertical y-axis from 0 to \$100 billion. The Year is plotted on the horizontal x-axis from 2010 to 2019. Different media are encoded by different colors: TV (light blue), Print (dark blue), Mobile (light purple), and All digital (dark purple).

Level 2 The ad spending for TV has risen gradually from around 60 billion to 75 billion from 2010 to 2019. This was the smallest increase of the media types that increased. All Digital ad spending increased rapidly from 25 billion to around 90 billion in this time period.

Level 3 Initially, Mobile ad spending increased steadily, but from 2014 it increased rapidly until 2019. By contrast, Print ad spending has been steadily decreasing. TV ad spending has been only gradually increasing from 2010 to 2019. By far, ad spending in the All Digital category has increased most dramatically, from almost \$0 in 2010 to around \$95 billion in 2019.

Level 4 From the ad spending trends presented in this visualization, it can be concluded that Digital media (such as TV advertising and especially Mobile) has become much more popular in recent years, and is now taking over traditional Print media.

Attention Check Walmart’s sales seem to be going up almost every year. Walmart’s highest year for sales was 2016. Walmart’s lowest year for sales was 2011.



#### Question Prompt for Blind Readers

Suppose that you are reading a business magazine article about the consumer price index inflation in the United Kingdom. You come across a visualization that plots the inflation rate for different bundles of goods. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

#### Ranking Choices

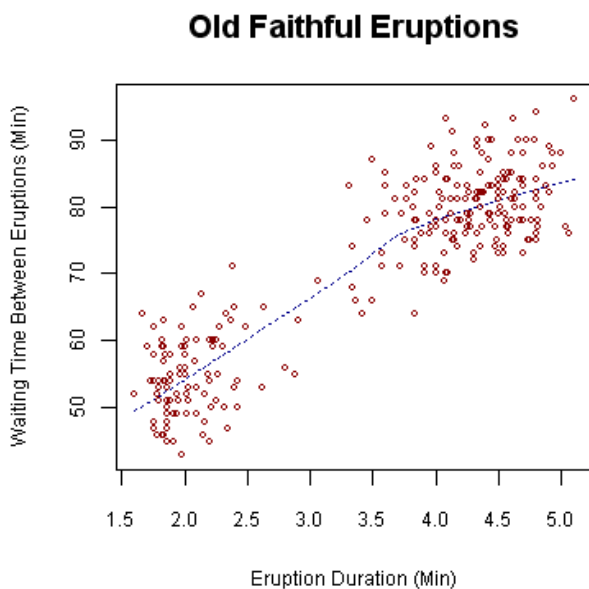
Level 1 This is a multi-line chart entitled "UK consumer price index inflation, annual" which shows the percent consumer price index inflation by year. The percent inflation is plotted on the vertical y-axis from 0 to 6%. The year is plotted on the horizontal x-axis from 2006 to 2015. The purple line represents "Overall" inflation rate for all goods. The blue line represents the inflation rate for goods "Excluding food, energy, alcohol, and tobacco."

Level 2 The "Overall" inflation rate reaches its highest point at the end of both 2008 and 2011. The inflation rate "Excluding food, energy, alcohol and tobacco" reached its highest point in mid 2011. The "Overall" inflation rate reached its lowest point in 2015. The inflation rate "Excluding food, energy, alcohol and tobacco" reached its lowest point in 2015.

Level 3 The "Overall" inflation rate and the inflation rate excluding food, energy, alcohol, and tobacco have remained fairly consistent over time. However, the "Overall" inflation rate greatly exceeds the "Excluded" rate in 2008 to 2009, suggesting that the inflation for food, energy, alcohol, and tobacco increased significantly during these years.

Level 4 The global economic recession of 2008 likely explains the inflation spike due to food, energy, alcohol, and tobacco. These goods are either a necessity or forms of escapism that can be habit forming. Thus, inflation can remain relatively stable because people will still buy these goods because they are addicted to them, or need them in order to survive or cope with hardships.

Attention Check Employment in the auto-industry decreased significantly in early 2009 (likely due to the financial crisis). Employment then rebounded to higher-than-before after the restructuring of General Motors and Chrysler.



#### Question Prompt for Blind Readers

Suppose that you are reading a popular science article about the eruptions at the Old Faithful geyser at the Yellowstone National Park in Wyoming, USA. You come across a visualization that is inaccessible. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

#### Ranking Choices

Level 1 This is a scatter plot entitled "Old Faithful Eruptions" that plots Waiting Time Between Eruptions (Min) by Eruption Duration (Min). The Waiting Time Between Eruptions is plotted on the vertical y-axis from 50 to 90 with an increment of 10 minutes. The Eruption Duration is plotted on the horizontal x-axis from 1.5 to 5.0 with an increment of 0.5 minutes.

Level 2 Many Old Faithful eruptions last from 4.0 to 5.0 minutes and the waiting time between them is 70 to 90 minutes. Also, many eruptions last from 1.5 to 2.5 minutes and the waiting time between them is 50 to 60 minutes.

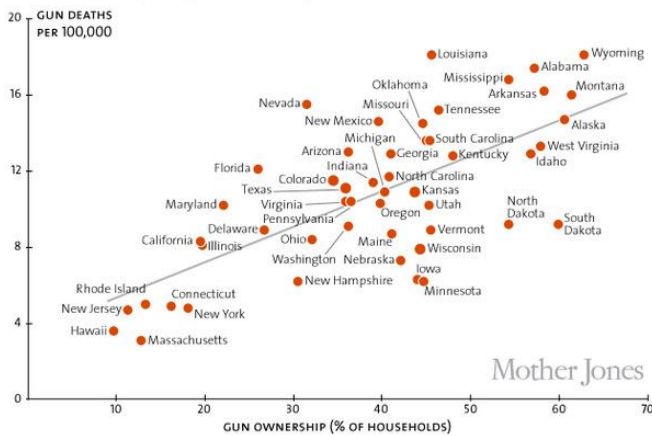
Level 3 Generally, the longer the waiting time between Old Faithful eruptions the longer that eruption will last. There are many very short eruptions with very short waiting time, and there are many very long eruptions with very long waiting time, but there are not many eruptions in-between (i.e., eruptions lasting around 3 minutes, with a waiting time of 70 minutes). Most eruptions either have a waiting time around 50 minutes or 80 minutes, and have a duration around 2.0 or 4.5 minutes, respectively.

Level 4 When Old Faithful eruptions have a short duration, it tends to be because the waiting time between them was short. This suggests that longer eruptions are due to a build-up of pressure in the geyser over time. Thus, the longer the waiting time between eruptions, the longer the eruption.

Attention Check The higher your MCAT and GPA are, the more likely you are to be accepted to Columbia University Medical School.



## Gun ownership vs. gun deaths, by state



## Question Prompt for Blind Readers

Suppose that you are reading a newspaper article about the rise in gun violence in the United States, and you encounter a visualization about the relationship between gun ownership and gun deaths, by state. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

## Ranking Choices

Level 1 A scatter plot entitled "Gun ownership vs. gun deaths, by state" that plots Gun Deaths Per 100,000 by Gun Ownership (% of Households). The Gun Deaths Per 100,000 is plotted on the vertical y-axis from 0 to 20. The Gun Ownership (by % of Households) is plotted on the horizontal x-axis from 0 to 70. Each data point represents a US state and is labeled with the name of the state.

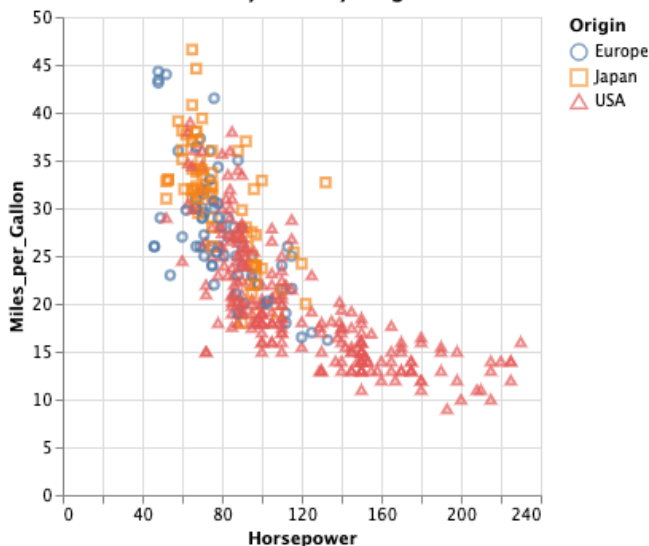
Level 2 There is a positive correlation between the percent of households that have guns and gun related deaths. The state with the highest gun ownership and corresponding gun deaths is Wyoming. The state with the lowest gun ownership and gun deaths is Hawaii. The most gun deaths per 100,000 people occurred in Louisiana and Wyoming, but Louisiana has 50% household gun ownership and Wyoming has 70% household gun ownership.

Level 3 The visual trend suggests that higher rates of gun ownership by household lead to more gun deaths. However, there are some exceptions to this trend, such as South Dakota, where 60% of households own guns, but there are only around 9 deaths per 100,000 people. Most states are clustered in the center of the x-axis (which plots gun ownership as percent of households) with an estimated 40% of households owning guns.

Level 4 Apart from Hawaii, the states with the lowest gun ownership (and lowest gun deaths) are geographically on the East Coast of the United States. The states with the highest gun ownership (and highest gun deaths) are geographically in the center of the United States. This geographic difference corresponds to the areas of the country that are more liberal and more conservative, respectively.

Attention Check Products in Family 1 appear to be the least sold overall. The product with the least units sold is from product Family 1. The product with the most units sold is from product Family 4. There appears to be a clear positive correlation between units sold and revenue in millions, therefore suggesting that all products have similar prices.

## Cars by Country Origin



## Question Prompt for Blind Readers

Suppose that you are reading a Consumer Reports article on the fuel efficiency and speed of cars produced in different countries (Europe, Japan, and the USA). You encounter a visualization that is inaccessible. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

## Ranking Choices

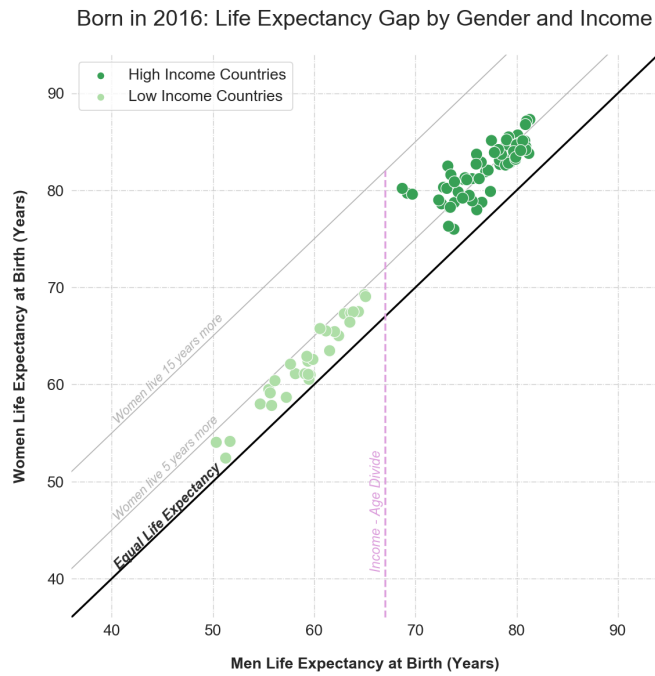
Level 1 This is a scatter plot entitled "Cars by Country Origin". The chart plots cars' Horsepower versus Miles\_per\_Gallon by Origin. Car Origin includes Europe, Japan, and the USA, encoded by Color and Shape, as follows: Europe (blue, circle), Japan (orange, square), and USA (red, triangle). Horsepower is plotted on the horizontal x-axis from 0 to 250 with an increment of 50. Miles\_per\_Gallon is plotted on the vertical y-axis from 0 to 50 with an increment of 10.

Level 2 The USA produces more cars than Europe or Japan. Japan produces the car with the greatest Miles\_per\_Gallon. The USA produces the car with the fewest Miles\_per\_Gallon. The USA produces the car with the greatest Horsepower. Europe produces the car with the least Horsepower.

Level 3 The overall trend suggests that the lower a car's Miles per Gallon, the higher a car's Horsepower, and vice-versa. Overall, cars produced in the USA have higher Horsepower, but consume more fuel than cars produced in Europe or Japan. European and Japanese cars tend to have higher Miles per Gallon, but they have lower Horsepower. In other words, European and Japanese cars are more efficient and consume less fuel, but they are not as fast as cars produced in the USA.

Level 4 Europe and Japan appear to prefer cars with higher Miles per Gallon, which is economically more practical and efficient. The USA produces a wide range of cars, which may appeal to a wider range of customers. Cars with higher Horsepower but lower Miles per Gallon tend to be luxury cars, so wealthy people who are car enthusiasts may be more attracted to the USA car market.

Attention Check 35% of people in China agree that their government drives most of their innovation. This is more than in any other country.



Source: World Bank

Question Prompt for Blind Readers

Suppose that you are reading an academic paper about how life expectancy differs for people of different genders and income. You encounter a visualization that is inaccessible via screen reader. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

Ranking Choices

Level 1 This is a scatter plot entitled "Born in 2016: Life Expectancy Gap by Gender and Income" that plots Women Life Expectancy at Birth (Years) by Men Life Expectancy at Birth (Years). The Women Life Expectancy at Birth is plotted on the vertical y-axis from 40 to 90 years. The Men Life Expectancy at Birth is plotted on the horizontal x-axis from 40 to 90 years. High Income Countries are plotted in dark green. Low Income Countries are plotted in light green. A 45 degree line from the origin represents Equal Life Expectancy between women and men.

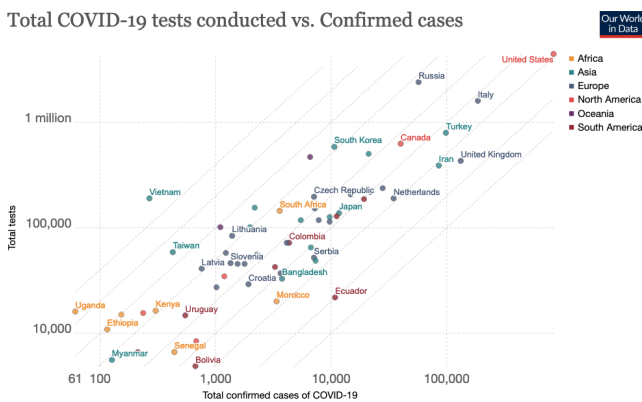
Level 2 In low income countries, men have an average life expectancy about 60 years, and women have an average life expectancy of about 65 years. In high income countries, the average life expectancy is about 77 years for men and 82 years for women.

Level 3 Overall, women have a slightly higher life expectancy than men. Women live around 5 to 10 years longer than men. Lower income countries have lower life expectancy than in high income countries. The low income countries are more scattered than the high income countries. There is a visible gap between high and low income countries, indicated by the Income-Age Divide line.

Level 4 People living in low-income countries tend to have a lower life expectancy than the people living in high-income countries. This is likely due to many societal factors, including access to healthcare, food, other resources, and overall quality of life. People who live in lower income countries are more likely to experience deprivation and poverty. They are more likely to have poor working and living conditions and can develop related health problems.

Attention Check A lot of rich countries are becoming poorer, whereas not many are becoming richer. A lot of poorer countries are catching up to the bigger countries, and very few are getting worse. This shows that overall countries are becoming more equal in terms of their GDP.

Total COVID-19 tests conducted vs. Confirmed cases



Our World In Data

Question Prompt for Blind Readers

Suppose that you are reading a newspaper article about the number of COVID-19 tests and cases in different countries and continents of the world. You encounter a visualization that is inaccessible via screen reader. The raw data presented in this visualization is not available to the reader. Which descriptive content below would be most useful to include in a textual description of the visualization?

Ranking Choices

Level 1 This is a scatter plot entitled "Total COVID-19 tests conducted vs. Confirmed cases" that plots the total tests by the total confirmed cases of COVID-19. The total tests are plotted on the vertical y-axis from 10,000 to 1 million. The total confirmed cases are plotted on the horizontal x-axis from 61 to 100,000 using a logarithmic scale. The plotted countries are color coded by their continent: Africa (gold), Asia (green), Europe (blue), North America (red), Oceania (purple), South America (burgundy).

Level 2 The number of COVID-19 tests ranges from 10,000 to 1 million. The number of confirmed cases ranges from 61 to 100,000. The USA has the highest number of tests and the highest number confirmed cases. Bolivia has the lowest number of tests and about 1000 confirmed cases. Uganda has a low number of tests and the lowest number of cases, about 100.

Level 3 In most continents, there is a strong positive correlation between the total number of tests conducted, and the number of COVID-19 cases: the higher the total of tests, the higher the number of cases, and vice-versa. The countries with the lowest number of tests and cases appear to be in Africa. In general, the countries with the highest number of tests and cases appear to be in North America and Europe.

Level 4 Overall, countries that are considered to be wealthier (such as the United States) have been able to conduct more tests and report more confirmed COVID-19 cases. Countries in Africa have done the fewest COVID-19 tests, but this can't confirm that they also have the lowest number of cases, which may be unknown due to their relatively limited testing.

Attention Check Poland, Romania and Lithuania have the largest positive change in GDP per capita. Greece has the largest negative change in GDP per capita. Countries that already have a high GDP per capita have lower change in GDP per capita.

### 3 EVALUATION QUESTIONNAIRE

See Paper Section 5.4 Qualitative Results.

Thank you for completing all the visualization description ranking tasks! We now ask you to complete the following questionnaire about your experience with data visualizations and with this study. Please review the following commitment to your data privacy.

- No personally identifying information will be revealed to anyone other than the researchers conducting this study.
- The records of this study will be kept private. In any sort of report we make public, we will not include any information that will make it possible to identify you.
- None of your responses will be linked to your email or name. The researchers will only review and analyze the data after it has been anonymized.

#### 3.1 Demographic Questions

What is your highest degree received?

- No degree
- High school diploma
- Bachelor's or Associate's degree
- Master's degree
- Doctorate degree
- Prefer not to say

What is your current age?

- 18-19
- 20-29
- 30-39
- 40-49
- 50-59
- 60-69
- 70-79
- 80+
- Prefer not to say

What gender(s) do you most identify with? Please write in the text field (or write "Prefer not to say").

Text input field

What is your field of work or study? Please write in the text field (or write "Prefer not to say").

Text input field

How long have you been working in your field?

- Less than a year
- 1-2 years
- 3-5 years
- 5-10 years
- 10-20 years
- More than 20 years
- Prefer not to say

What screen reader did you use while completing this study?

- JAWS
- NVDA
- VoiceOver
- Other

### 3.2 Visualization Questions

How often do you encounter data visualizations that are inaccessible to you?

- Everyday
- Multiple times a week
- At most once a week
- Once or twice a month
- Almost never

In which contexts do you most often encounter inaccessible data visualizations?

- Social media (Twitter, Facebook, etc.)
- Newspaper websites (The Wall Street Journal, The New York Times, Fox News, etc.)
- Educational materials (textbooks, audiobooks, etc.)
- Personal records (billing statements, banking information, etc.)
- I encounter inaccessible visualizations in all of these contexts

If you encounter inaccessible visualizations in any other contexts, please describe them below.

Text input field

Please rate your agreement with the following statement: I have a good understanding of data visualization concepts and terminology, such as “bar chart”, “scatter plot”, “x-axis”, “y-axis” and so on.

1. Strongly Disagree 2. Disagree 3. Somewhat Disagree 4. Neither Agree Nor Disagree 5. Somewhat Agree 6. Agree 7. Strongly Agree

Please rate your agreement with the following statement: Providing textual descriptions of data visualizations is important to me.

1. Strongly Disagree 2. Disagree 3. Somewhat Disagree 4. Neither Agree Nor Disagree 5. Somewhat Agree 6. Agree 7. Strongly Agree

Please rate your agreement with the following statement: I have a good understanding of statistical concepts and terminology, such as “mean”, “standard deviation”, “regression”, “outliers”, and so on.

1. Strongly Disagree 2. Disagree 3. Somewhat Disagree 4. Neither Agree Nor Disagree 5. Somewhat Agree 6. Agree 7. Strongly Agree

Please rate your agreement with the following statement: I often feel that important public information (such as the spread of COVID-19, or the status of political elections) is inaccessible to me, because it is only available in a visual format.

1. Strongly Disagree 2. Disagree 3. Somewhat Disagree 4. Neither Agree Nor Disagree 5. Somewhat Agree 6. Agree 7. Strongly Agree

Please rate your agreement with the following statement: I am familiar with existing visualization description guidelines (such as those from the World Wide Web Consortium’s Web Accessibility Initiative, or W3C’s guidelines for Digital Talking Books)?

1. Strongly Disagree 2. Disagree 3. Somewhat Disagree 4. Neither Agree Nor Disagree 5. Somewhat Agree 6. Agree 7. Strongly Agree

When reading an article via screen reader, I prefer that the information conveyed by a data visualization is presented in the following way(s).

- Presented as a standalone description, separate from the main text of the article.
- Presented within the main text of the article.
- Presented both as a standalone description and presented in the main text of the article.

In the text field below, please elaborate on what makes a textual description of a visualization useful to you? What makes a description more or less useful?

Text input field

In the text field below, please comment on your experience with data visualizations. Do you ever encounter barriers to reading data visualizations? If so, what are they? And in which contexts do you usually encounter these barriers?

Text input field

In the text field below, please leave any additional thoughts, comments, or suggestions about this study, or about data visualizations in general.

Text input field