Graphs and Gapminder

After completing this section, students should be able to:

- Read a static plot/graph of one to many variables and use it to tell a story
- Read a dynamic complex plot/graph of many variables and use it to tell a story
- Critique a graph and point to issues of axis truncation, axis compression, and misuse of area and volume.
Graphs

Type of graph: histogram: represents amount by ...
Distribution of annual household income in the United States
2010 estimate

Median household income was roughly $50,000.

These two groups include households reporting income greater than $200,000 (approximately 4 percent of households).

The top 25 percent reported income greater than $85,000.

The top 10 percent reported income greater than $135,000.

Categories in $5,000 increments with the exception of the last two groups

**Question.** About what percent of households earn between $30,000 and $35,000?

**Question.** About what percent of households earn between $90,000 and $115,000?

**Question.** On the plot of the US annual household income, are there more households with income under $5,000 a year or over $200,000 a year?

A. Substantially more households have incomes under $5,000 a year.
B. Substantially more households have incomes over $200,000 a year.
C. About the same number of households are in both categories.

**Question.** Are there more households with income below $25,000 or between $90,000 and $115,000?

**Question.** Why do the bars jump up higher at the far right?
**Question.** Ignoring the higher right two bars, is this graph symmetric – i.e., do the left and right sides look like mirror images?

**Question.** What does this shape tell you about rich and poor in the US?
Type of graph: scatterplot
1. What variables are represented in the graph?

2. What relationships exist between these variables?

3. How might you explain why these relationships exist?
Each group, pick one graph (or one related pair of graphs) from the following collection. For your graph, answer the following questions.

1. What variables are represented in the graph?

2. What relationships exist between these variables?

3. How might you explain why these relationships exist?
<table>
<thead>
<tr>
<th>State</th>
<th>DACA beneficiaries</th>
<th>Age at entry</th>
<th>Percent of DACA beneficiaries</th>
</tr>
</thead>
<tbody>
<tr>
<td>California</td>
<td>222,795</td>
<td>0</td>
<td>4.5%</td>
</tr>
<tr>
<td>Texas</td>
<td>124,300</td>
<td>1</td>
<td>7.6</td>
</tr>
<tr>
<td>Illinois</td>
<td>42,376</td>
<td>2</td>
<td>8.3</td>
</tr>
<tr>
<td>New York</td>
<td>41,970</td>
<td>3</td>
<td>9.7</td>
</tr>
<tr>
<td>Florida</td>
<td>32,795</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Arizona</td>
<td>27,865</td>
<td>5</td>
<td>8.7</td>
</tr>
<tr>
<td>North Carolina</td>
<td>27,385</td>
<td>6</td>
<td>7.4</td>
</tr>
<tr>
<td>Georgia</td>
<td>24,135</td>
<td>7</td>
<td>7.4</td>
</tr>
<tr>
<td>New Jersey</td>
<td>22,024</td>
<td>8</td>
<td>6.5</td>
</tr>
<tr>
<td>Washington</td>
<td>17,843</td>
<td>9</td>
<td>6.2</td>
</tr>
<tr>
<td>Colorado</td>
<td>17,258</td>
<td>10</td>
<td>5.6</td>
</tr>
<tr>
<td>Nevada</td>
<td>13,070</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Virginia</td>
<td>12,134</td>
<td>12</td>
<td>3.9</td>
</tr>
<tr>
<td>Oregon</td>
<td>11,281</td>
<td>13</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>14</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15</td>
<td>3.7</td>
</tr>
</tbody>
</table>
A slower field at this year’s Boston Marathon

Finish time for winners of the Boston Marathon, by country

Using unofficial times for 2018 winners.

FiveThirtyEight

SOURCE: BOSTON ATHLETIC ASSOCIATION
Outcomes for Shelter Dogs and Cats at Dallas Animal Services

- Adopted
- Returned to owner
- Transferred to rescue
- Released
- Lost or died
- Euthanized
Projected career earnings for college graduates in ...
By the time Hurricane Harvey left Texas, parts of Houston had been inundated by more than 50 inches of rain.
Difference between average winter temperatures by year and a 20th century average

Note: Contiguous U.S. only.
Bad Graphs

For each graph, find the weaknesses of the graph. Can you think of ways to improve it?
Group 1:

Average number of actual weekly hours of work in main job, full-time employees, 2013

- Romania: 41.2 hours
- UK: 40.8 hours
- Germany: 40.4 hours
- Poland: 39.9 hours
- Netherlands: 39.8 hours
- Slovak Republic: 39.8 hours
- Czech Republic: 39.7 hours
- EU-28 average: 39.5 hours
- Spain: 39.5 hours
- Hungary: 39.2 hours
- Sweden: 38.6 hours
- Belgium: 38.1 hours
- Italy: 38.0 hours
- France: 37.4 hours

Source: Eurofound 2014
Group 2:

Average Annual Global Temperature in Fahrenheit
1880-2015
Group 3:

**Gun deaths in Florida**

Number of murders committed using firearms

Source: Florida Department of Law Enforcement

C. Chan 16/02/2014
Group 4:

It’s Who You Know

By ages 35–44, people are likely to have the strongest network of fellow entrepreneurs, but these personal social networks begin to decline over time.

- 30%: 18-24
- 31%: 25-34
- 26%: 35-44
- 20%: 45-54
- 13%: 55-64
- 13%: 65+

Interacting with other entrepreneurs fosters possible role models, networking opportunities, advice, and encouragement.
Group 5:

**Graph A**

- **X-axis:** Mass shooters
- **Y-axis:** United States

**Graph B**

- **X-axis:** Guns per 100 people
- **Y-axis:** Mass shooters per 100 million people

Countries marked: Philippines, Russia, China, India, Afghanistan, Yemen, United States, France, Canada, Iraq.
Gapminder

Gapminder is a plotting tool that lets you explore relationships between variables by country.

- Go to www.gapminder.org/tools
- Click on the arrows by the axes labels to change the variables being plotted. You can also change from linear to logarithmic scales.
- Click the play button to see the change over time.
Find a good story to tell with a Gapminder graph.

• What variables would be of interest to explore their relationships?
• How is best to use Gapminder to explore these relationships?
• How might you explain why these relationships exist?
Homework

1. Find a graph online or in the media that you think is misleading the viewer.
   (a) Include an image of the graph and an informal citation of where you found it.
   (b) How is the graph misleading the viewer? How would you fix it?
   (c) Do you think it is intentionally manipulative? Why or why not.

2. Go to Gapminder at https://www.gapminder.org/tools/ and watch the How To Use short video (link at the top).
   (a) Change the x- and y-axes to variables that have a relationship that you find interesting. Include a screenshot of your graph (you can pick which year to stop the video on).
   (b) Describe the relationship between the variables. Specifically, what happens to the y-axis variable as the x-axis variable increases, and what happens over time?
   (c) Provide a potential explanation for your observations in part (b).
   (d) Describe something you find interesting about this graph.

3. Look at the graph about ride-hailing services below from NYTimes.com. The data was generated by RideAustin, a nonprofit ride-hailing service in Austin, Texas.
Each point in this graph represents one of the service’s 1.5 million ride-hailing trips over the course of 10 months.

(a) What variables are related in this graph?

(b) Describe the relationship between the variables, by specifying what happens to one variable as another variable increases.

(c) Why might this relationships exist?
(d) Why are there no data points in the upper right quadrant of the graph?

(e) According to the NY Times, a researcher analyzed the data in this graph to find that the optimal number of ride requests per driver per hour was 3.4. What about the shape of the graph makes it seem plausible that 3.4 ride requests per hour is most efficient for drivers and riders?