Data Visualization
Learning Goals

You should be able to:

• [CT Application] Be able to assess strengths and weaknesses of simple data visualizations
• [CT Building Block] List factors that you should consider when interpreting an infographic or statistic
• [CT Building Block] Understand the steps involved in creating data visualizations
Why data visualization?

Visual representations of data can help us to

- Communicate known results
- Explore data to uncover patterns
- Carry out tasks more efficiently
- Make sound policy decisions
Data visualization and statistics

- Many visualizations convey statistical information.
- Some statistics you may be familiar with are:
  - Mean/Average (sum of numbers, divided by the total number of numbers)
  - Median (the number in the middle of a set of numbers when the numbers are arranged in order)
  - Mode (the number that occurs most frequently)

Visit [this CDC lesson](https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson2/section8.html) for more information on statistics.
Statistics can lie!

- Numbers can be manipulated to skew a story
- You should check who is producing or promoting a statistic
  - Are they directly related?
  - Do they have an agenda they want to push?
  - **How** did the number come about?
When looking at data visualizations...

- A lot of the same rules apply!
- You want to look at
  - who the visualization is from
  - what kind of agenda they may have
  - how the numbers they are using came about
- Sometimes, visualizations will take advantage of people’s first impressions...

https://xkcd.com/1138/
Data visualizations can be static
Data visualizations can be dynamic
Cornell Lab of Ornithology: Bird migration patterns

https://www.allaboutbirds.org/mesmerizing
migration-watch-118-bird-species-migrate-across-a-map-of-the-western-hemisphere/
Data visualizations can be dynamic

“By using eBird data and other forms of migration tracking information, we’re getting a more detailed picture than ever before about where and when birds migrate. That’s the kind of information we need to make smart conservation decisions for species that live in vastly different regions during the year…”
Creating a Visualization

Determine your domain

Break down your data/task

Select the appropriate visual encodings/interaction

Create an algorithm for the computer to create the visualization
Determining your domain

- **domain** is an specific area (e.g., microbiology, chemistry, economics, etc.)
- It has its own vocabulary/terminology to describe things
- There are often existing conventions for describing or working with data
- There may also be tools associated with these tasks
- It is important to know about the domain you’re working with!
Determining your domain

- At the end of this stage, you should know what your user wants and needs from your visualization.
- In this stage, you may use methods like:
  - Interviewing the users
  - Observing what the users normally do
  - Doing some research into the users your visualization is trying to target.

Example: Describe your domain

- The domain for the data provided is educational information.
- The predominant tools that people currently use to work with this kind of data are spreadsheets (e.g., Excel) or R (a statistics programming language).
Breaking down your data/task

- At this point, you have a specific task but this task is probably in very domain specific language.
- You can extrapolate this task into more general terms. This can help you determine what visual encodings/interaction (next step) will be needed!
Breaking down your task

Break your task down into three steps:

- **Analyze**: What is the purpose of your visualization? To consume existing data or to produce new data?
- **Search**: Do we know what we want to find? Do we know where to find whatever it is we are looking for?
- **Query**: How do you want to look at the data once you find it? Do you want to look at one data point, multiple data points, or all data points?
### Breaking down your data/task: Search

#### Search

<table>
<thead>
<tr>
<th>Location known</th>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location known known</td>
<td><img src="lookup.png" alt="Diagram" /></td>
<td><img src="browse.png" alt="Diagram" /></td>
</tr>
<tr>
<td>Location unknown</td>
<td><img src="locate.png" alt="Diagram" /></td>
<td><img src="explore.png" alt="Diagram" /></td>
</tr>
</tbody>
</table>
Breaking down your data/task: Query

- Query
  - Identify
  - Compare
  - Summarize
Example: Break down your task

- Our task is:
  “..to create something that will let people see how the percentage of women in different majors has changed over time”
- Although we say we are creating something, we are actually trying to help people look at data
- Since we aren’t producing any new numbers from our existing data, we are not producing numbers—we are consuming them
Example: Using not enough idioms for something can make a visualization hard to read / confusing.

- Computer Science: 13.6%
- Engineering: 0.8%
- Psychology: 44.4%
- Health Professions: 77.11%
- Social Sciences: 36.8%
Example: Using not enough idioms for something can make a visualization hard to read / confusing
Example: Certain idioms (or certain implementations of idioms) are better.
Create the visualization using a computer tool

- After making all these decisions, we can start creating the visualization on a computer.
- There are many nice visualization tools available now for this purpose – you can play with some in lab.

In lab you’ll get to play with visualizations and do peer evaluations of each others’
Great overview of a presentation by Tamara on the blog of Dr. Jenny Cham: http://jennycham.co.uk/2014/07/a-peek-into-the-world-of-data-visualisation-with-prof-munzner/
Learning Goals Revisited

You should be able to:

• [CT Application] Be able to assess strengths and weaknesses of simple data visualizations

• [CT Building Block] List factors that you should consider when interpreting an infographic or statistic

• [CT Building Block] Understand the steps involved in creating data visualizations