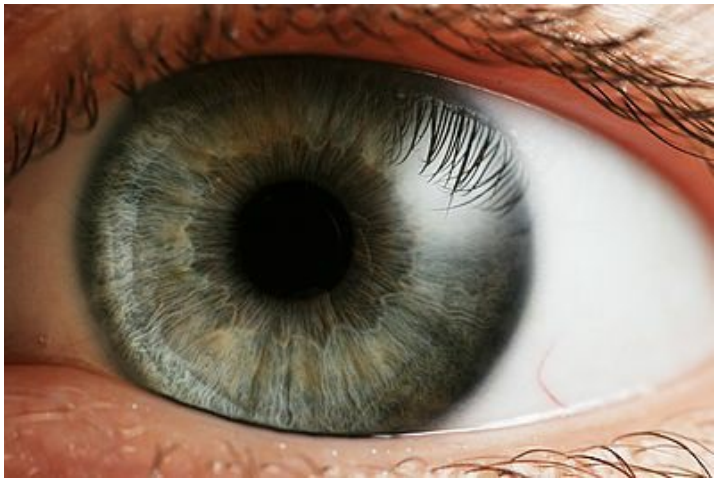
PLTW COMPUTER SCIENCE

Activity 3.1.4

Pie Charts and Bar Charts

Introduction

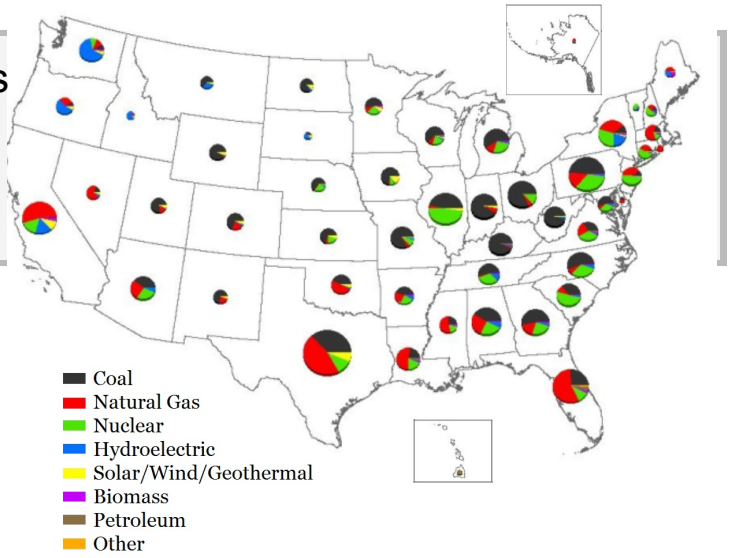
To find patterns in data, you need tools. Data visualizations are the basic tools of a data scientist. Computers can create a single graphic from millions of data points. The human brain is very good at abstracting patterns from pictures. Consider the following three quotes or sayings.



- *"Exploratory data analysis" is a state of flexibility, a willingness to look for those things that we believe are not there, as well as those we believe to be there.* (John Tukey, founder of this new side of data science)
- *A picture says a thousand words.*
- *Seeing is believing.*

> Refer to your downloadable resources for this material. Interactive content may not be available in the PDF edition of this course.

What do these three ideas have in common? What kinds of pictures can change our beliefs? Look at the data picture in the introduction titled "Electricity generation by source and state in 2010". What patterns do you see?



Electricity generation by source and state in 2010

Materials

- Computer with Enthought Canopy distribution of *Python*® and access to Internet
- A spreadsheet program, such as Microsoft® Excel®

Procedure

Part I: Pie Charts

1. Form pairs as directed by your teacher. Meet or greet each other to practice professional skills. Set team norms.
2. Refer to your downloadable resources for this material. Interactive content may not be available in the PDF edition of this course.

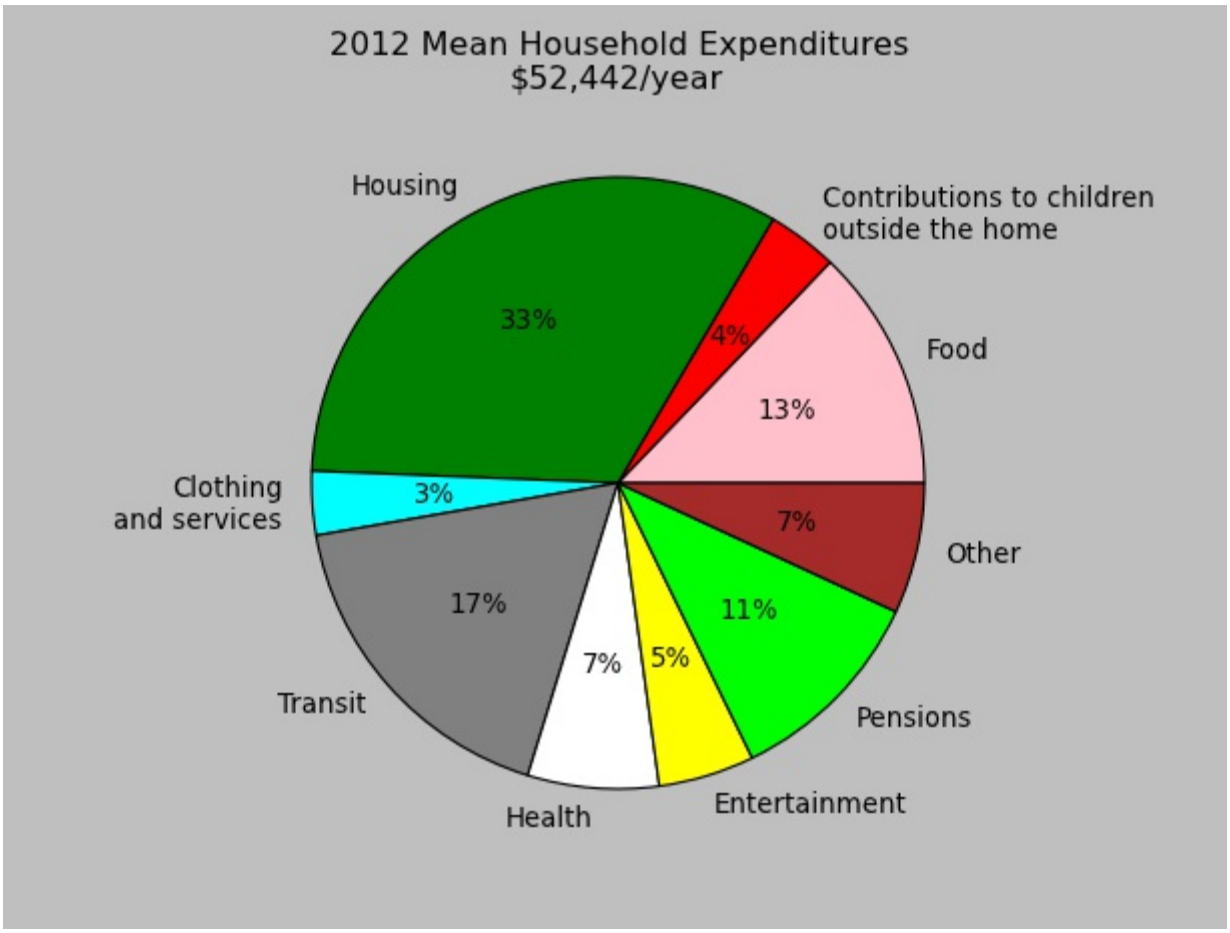
Pie charts are useful for displaying how 100% of a quantity is broken down based on categories. They should obey these guidelines.

- The full circle must represent 100% of the quantity.
- Two variables are needed to make a pie chart:
  - the labels of the categories
  - the quantity in each category

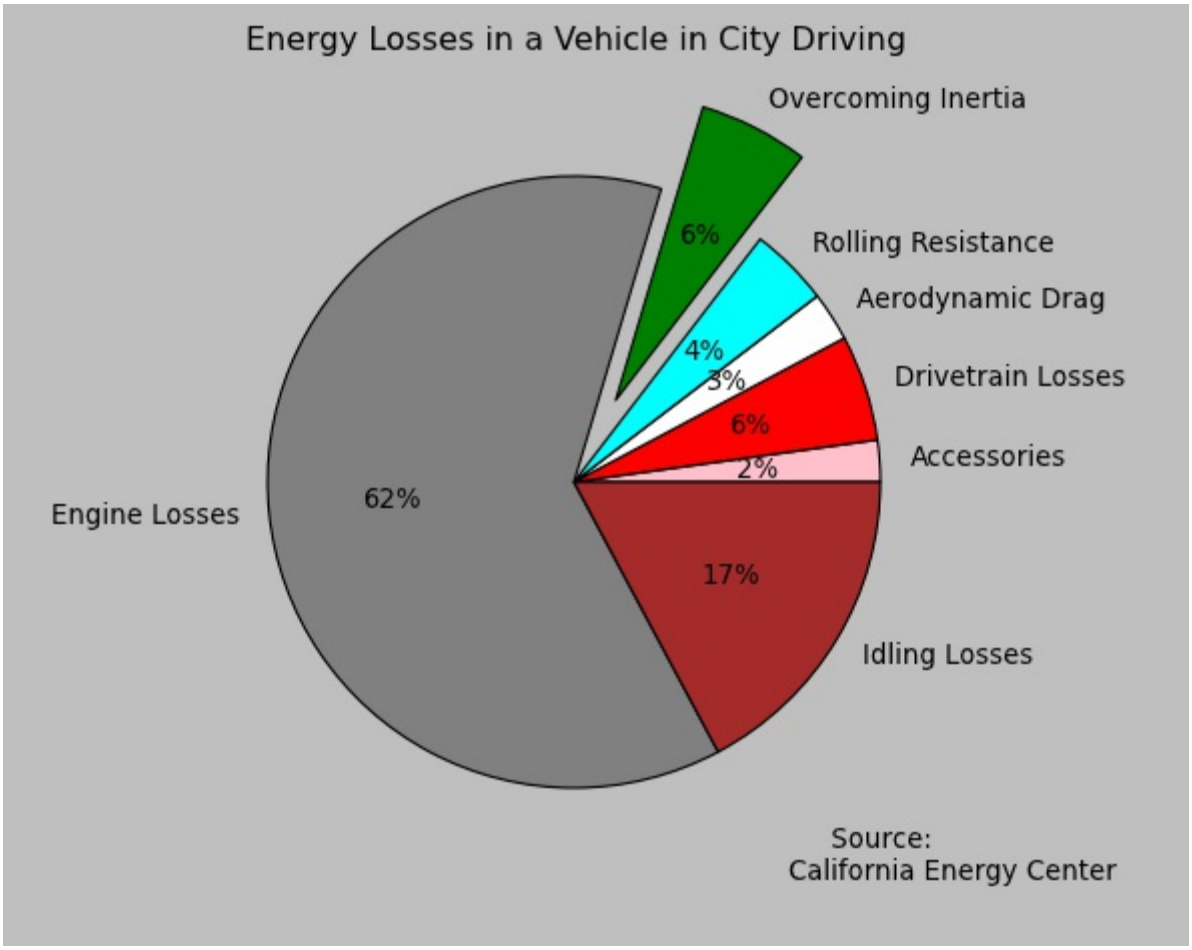
Pie charts are most useful for quickly scanning the relative composition of something with 5 or fewer categories, especially across multiple instances of the 100%, such as in the graphic in the introduction.

Critique the following two visualizations.

- Suggest improvements for this visualization.



- Suggest improvements for this visualization.



3. To create a pie chart in *Python*, place the category labels in one list and the quantities in another list. Call the `pie()` method on a `SubplotAxes` object as shown below. The `pie()` method is documented at **http://matplotlib.org/1.2.1/api/axes_api.html?highlight=pie#matplotlib.axes.Axes.pie**
    - One of the following two links contains some data appropriate for a pie chart. Examine the data and decide which link contains the data appropriate for a pie chart. Explain your reasoning.

      http://www.bls.gov/news.release/empsit.t13.htm

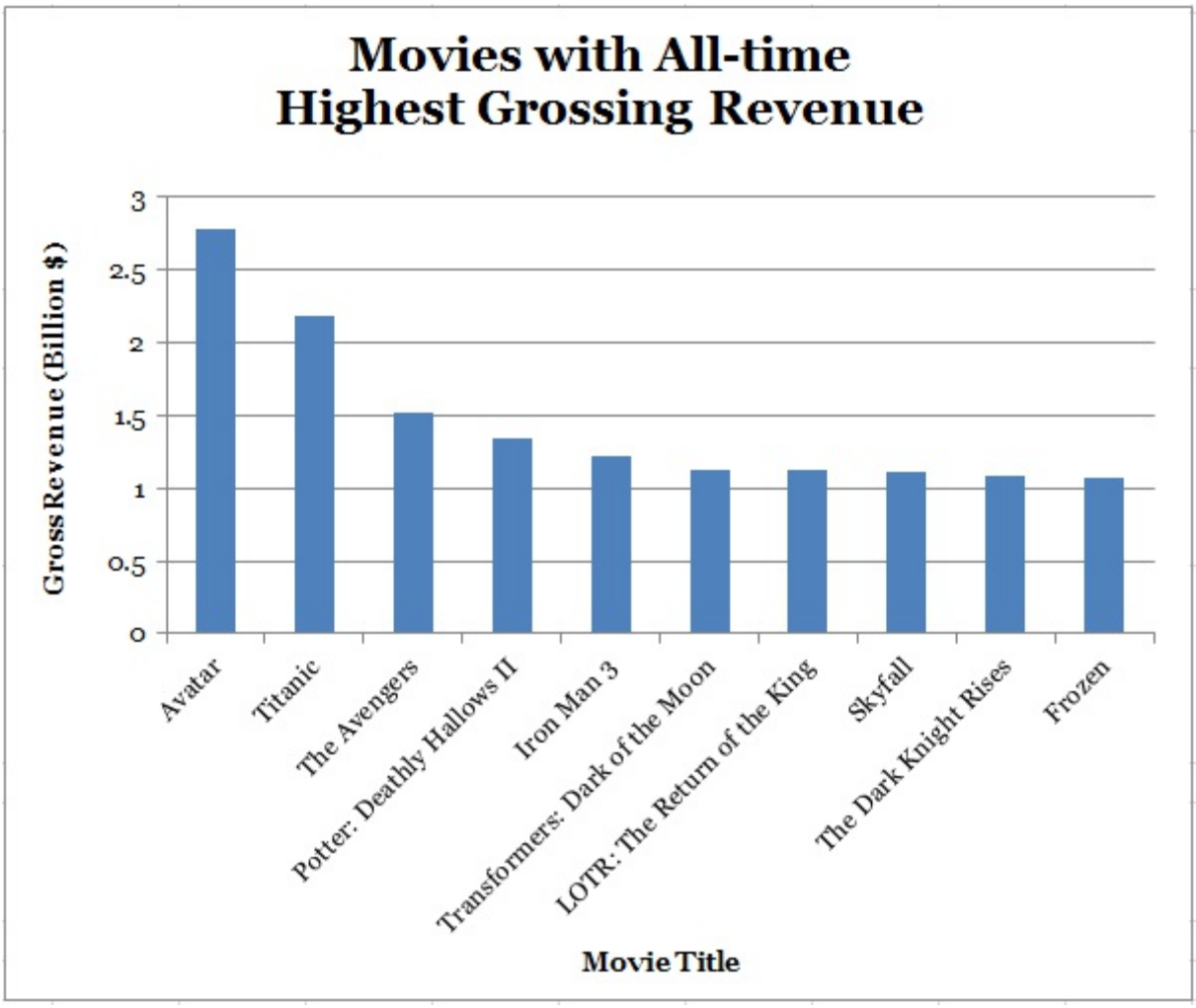
      http://www.bls.gov/news.release/empsit.t19.htm

    - Use *Python* to create a pie chart using matplotlib based on a group of data from the link you chose. Hint: use either subtotals or their components, but do not use both. You may use data about a different topic if approved by your teacher.

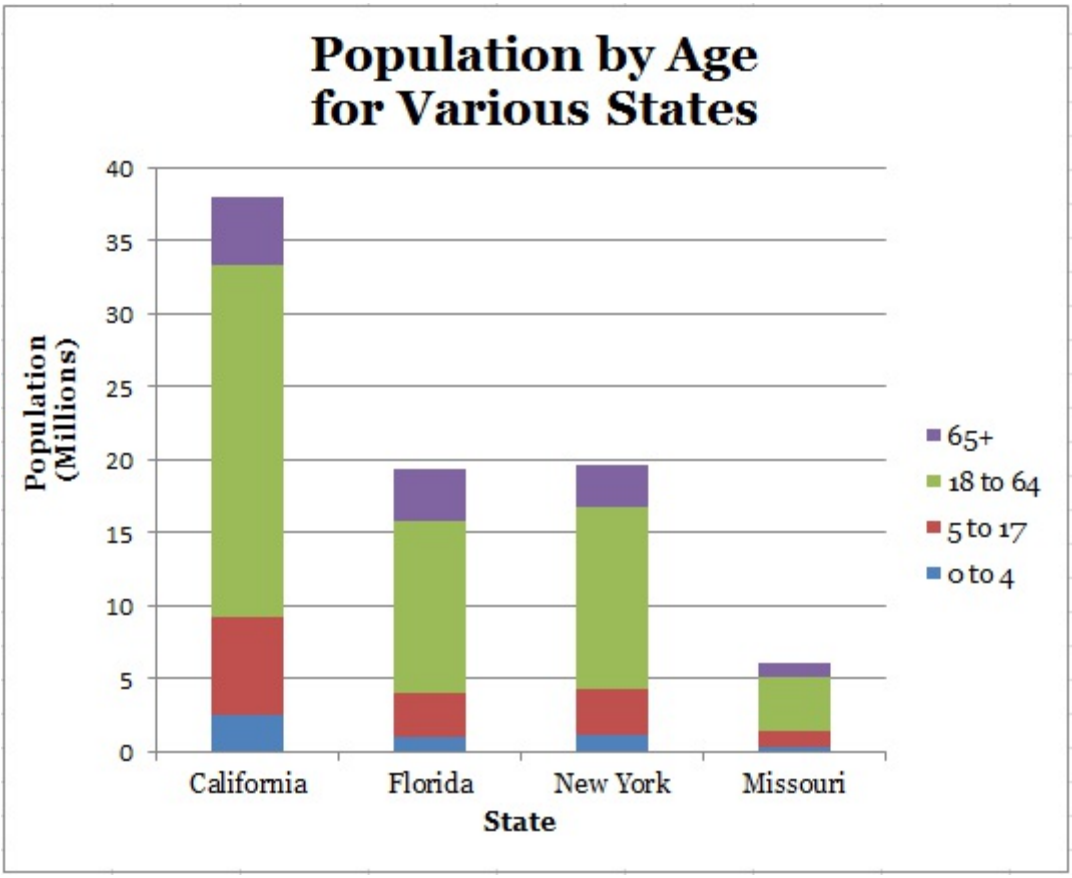
Part II: Bar Graphs

4. Pie charts emphasize the comparison of quantities that add up to a meaningful 100%. If the quantities being compared do not add up to a meaningful total, use a **bar graph** instead. A stacked **bar graph** is useful for comparing a total quantity for different categories while also comparing the composition of those totals. However, the stacked bar graph makes it difficult to compare the components; for such comparison a **side-by-side bar graph** should be used.

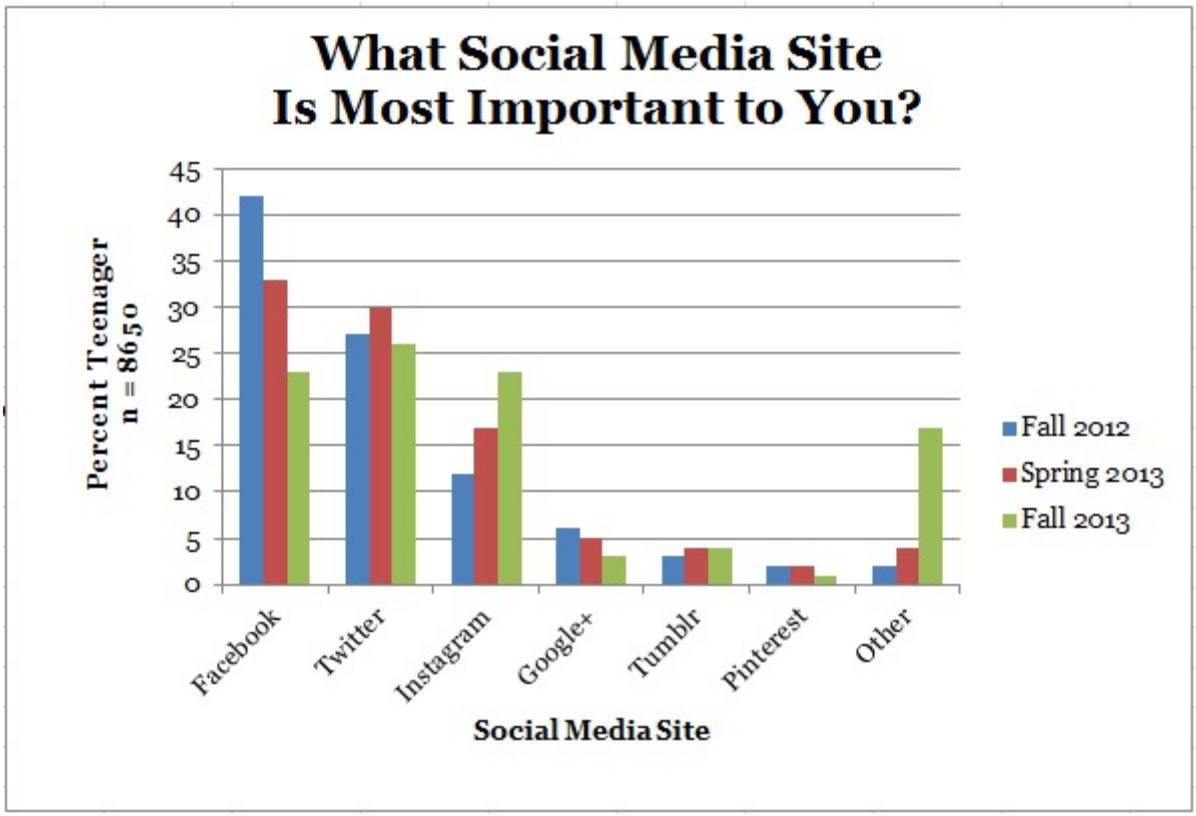
- Explain why the data in this graphic should not be represented by a pie chart. What pattern do you observe in this data?



- This data could have been visualized with four pie charts or with side-by-side bar graph. Identify the advantages of each of the three methods.



- This data could have been represented as 3 pie charts or as 7 time series. Identify the advantages of each of the three methods.



5. Use Excel to create a visualization from Census data using the following steps.

- Visit https://www.census.gov/quickfacts/table.
- Enter a zip code, city, county into the search bar. When you see your desired entry, click on the entry.
- Select the **More** ellipse (...) icon near the top right of the page, and select the **Excel** format to download.
- Select another county or city and repeat steps b and c two or more times.
- Open your downloaded files in Excel.
- Copy the data from the various counties or cities into a single spreadsheet, taking care to align data of the same kind within a single row. Delete rows with "X" or "NA" as data and delete duplicate columns.
- Select one row or group of rows to compare among the counties or cities you selected. Create a group of cells in the spreadsheet that has only the relevant data, with rows and columns labeled.
- Highlight only the relevant data and select **Insert** > (Charts panel) **Column** and select the appropriate 2-D column graph.
- Add an x-axis title, a y-axis title, and a chart title. Remove or add a legend as appropriate. In Office 2010, these are in the Chart Layout tool ribbon. For other productivity software, use Google to identify the method for adding these titles.

Conclusion

1. Data visualization is a creative tool.
    - Humans use data visualization to explore and understand data to discover unknown patterns. Explain why data visualization is important for presenting data from computer to human.
    - Data visualization allows one person who has discovered patterns in data to communicate those patterns to another person. Explain why data visualization is important for communicating from human to human.
    - When creating a data visualization, creativity plays an important role. Describe why the role of creativity might be different depending on whether the purpose is exploration/discovery or human-to-human communication.
2. How do you decide what type of data visualization to use?