CompSci 190: Tables & Graphs

Jeff Forbes
January 24, 2019
Plan For Today (PFTW)

• Create tables from datafiles and other sources
• Consider different methods for visualizations of data
• Start working on HW 1
Minard’s Visualization

- Napoleon’s 1812 invasion of Russia
  - # of soldiers
  - direction of the march
  - latitude and longitude of each city
  - temperature on the return journey
  - dates
Data in a tabular form

- What information does this visualization hide? What is more clear?

<table>
<thead>
<tr>
<th>Longitude</th>
<th>Latitude</th>
<th>City</th>
<th>Direction</th>
<th>Survivors</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>54.8</td>
<td>Smolensk</td>
<td>Advance</td>
<td>145000</td>
</tr>
<tr>
<td>33.2</td>
<td>54.9</td>
<td>Dorogobouge</td>
<td>Advance</td>
<td>140000</td>
</tr>
<tr>
<td>34.4</td>
<td>55.5</td>
<td>Chjat</td>
<td>Advance</td>
<td>127100</td>
</tr>
<tr>
<td>37.6</td>
<td>55.8</td>
<td>Moscou</td>
<td>Advance</td>
<td>100000</td>
</tr>
<tr>
<td>34.3</td>
<td>55.2</td>
<td>Wixma</td>
<td>Retreat</td>
<td>55000</td>
</tr>
<tr>
<td>32</td>
<td>54.6</td>
<td>Smolensk</td>
<td>Retreat</td>
<td>24000</td>
</tr>
<tr>
<td>30.4</td>
<td>54.4</td>
<td>Orscha</td>
<td>Retreat</td>
<td>20000</td>
</tr>
<tr>
<td>26.8</td>
<td>54.3</td>
<td>Moiodenoxo</td>
<td>Retreat</td>
<td>12000</td>
</tr>
</tbody>
</table>

**float:** decimal number

**string:** text

**int:** integer
Table Operations

- **t.select(label)** - constructs a new table with just the specified columns
- **t.drop(label)** - constructs a new table without the specified columns
- **t.sort(label)** - constructs a new table, with rows sorted by the specified column
- **t.where(label, condition)** - constructs a new table with just the rows that match the condition
Table Methods

- Creating and extending tables:
  - `Table().with_columns` and `Table.read_table`
- Finding the size: `t.num_rows` and `t.num_columns`
- Referring to columns: labels, relabeling, and indices
  - `t.labels` and `t.relabeled`; column indices start at 0
- Accessing data in a column
  - `t.column` takes a label or index and returns an array
- Using array methods to work with data in columns
  - `a.item(row_index)` returns a value in an array
  - `a.sum()`, `a.min()`, `a.max()` or `sum(a)`, `min(a)`, `max(a)`
Manipulating Rows

- \texttt{t.sort(column)} sorts the rows in increasing order
- \texttt{t.take(row\_numbers)} keeps the numbered rows
  - Each row has an index, starting at 0
- \texttt{t.where(column, are\_condition)} keeps all rows for which a column's value satisfies a condition
- \texttt{t.where(column, value)} keeps all rows for which a column's value equals some particular value
- \texttt{t.with\_row} makes a new table that has another row
Decennial Census

• Count how many people are in the US
• U.S. Constitution. Article I, Section 2

"Representatives and direct Taxes shall be apportioned among the several States which may be included within this Union, according to their respective Numbers . . . The actual Enumeration shall be made within three Years after the first Meeting of the Congress of the United States, and within every subsequent Term of ten Years, in such Manner as they shall by Law direct."

• Census Bureau estimates the population in intervening years
Census Table

• Values have column-dependent interpretations
  ○ The SEX column: 1 is Male, 2 is Female
  ○ The POPESTIMATE2010 column: 7/1/2010 estimate

• In this table, some rows are sums of other rows
  ○ The SEX column: 0 is Total (of Male + Female)
  ○ The AGE column: 999 is Total of all ages

• Numeric codes are often used for storage efficiency. Why?

• Values in a column have the same type, but are not necessarily comparable (AGE 12 vs AGE 999)

The rate of increase per unit time

After one time unit, a quantity $x$ growing at rate $g$ will be
$$x \times (1 + g)$$

After $t$ time units, a quantity $x$ growing at rate $g$ will be
$$x \times (1 + g)^t$$

If after and before are measurements of the same quantity taken $t$ time units apart, then the growth rate is
$$\left(\frac{\text{after}}{\text{before}}\right)^{\frac{1}{t}} - 1$$

What is the annualized growth rate of the population?
What’s next?

• Read Chapter 6-7 of *Computational and Inferential Thinking*

• Start working on Homework 1