Spreadsheets and Basic Data Operations, and Basic Visualizations
CS102 - Apr 6 Thu

A huge fraction of the world's structured data is managed and manipulated in spreadsheets. Excel is the dominant tool, but we'll use Google Sheets which is also very powerful.

Students should work along on their computers.

Basic spreadsheet operations

- **Importing/exporting from/to files: CSV, TSV**
  File > Import > Upload > Select YelpRestaurantsSample.csv, YelpReviewsSample.csv
  Note: exporting in CSV/TSV saves values, not formulas
- **Inserting (deleting) rows/columns**
  Right-click column header (**A**) > delete column
  Right-click row header (**1**) > delete row
- **Formulas**
  [YelpReviewsSample.csv] Converting Yelp 5-Star rating to 10-Point scale
  10-Point = A2*2
  To apply formula to whole column, double click bottom right corner of cell or drag bottom right corner to the bottom.

Basic data operations

- **Sorting**
  Sort reviews by star rating in [YelpReviewsSample.csv]
  > Select all the data (crtl/cmd + A)
  > Filter/Funnel Icon
  > Click column name icon
  > Sort A to Z
- **Hiding columns**
  Right-click > Hide Column
- **Freezing rows**
  Select row > Click View > Freeze > 1 row
- **Filtering rows**
  Filter all reviews greater than 2 in [YelpReviewsSample.csv]
  > Select all the data (crtl/cmd + A)
  > Filter/Funnel Icon
  > Click column "stars" icon
  > Filter by condition
  > Select "Greater than"
  > Type "2" into *value*
  > Click funnel icon again to exit filter view
  Annoyance: no way to save filtered data except copy-paste
- **Aggregation**
  Average Rating =average(A2:A1419)
  Max rating = max(A2:A1419)
- **Aggregation with Condition**
  Number of ratings above 3 Stars =COUNTIF(A2:A1419,">3")

Average Rating for ratings >3 stars =AVERAGEIF(A2:A1419,">3")

- **Grouped Aggregation**
  Calculate Average Rating by Restaurant
  > At Cell D2 insert formula =UNIQUE(B2:B1419)
  > =AVERAGEIF(B$2:B$1419,D2,A$2:A$1419)
  >Range I want to search over: (Sushi House, Sushi House, La Piazza, …)
  >What I want my search range to match with: (Sushi House)
  >Values you want to average over: (5,4,4,...)
  > Apply formula to rest of column (don't forget the dollar signs)
- **Absolute references (Dollar Sign $)**
  $A$2 **-** The column and row do not change when copied
  A$2 - The row does not change when copied
  $A2 - The column does not change when copied
- **Joining**
  > Copy grouped aggregation results over to [YelpRestaurantsSample] sheet columns I+J
  > In Cell F2 insert formula =VLOOKUP(D2,$I$2:$J$47,2,false)
  >What I want to match with
  >Range to lookup over (grouped aggregation of restaurants and their ratings)
  >Index of the column (rating) of the range that we are looking up that we want to retrieve. In this case, it is 2nd from the left, hence 2.
  >Leave as False
  > Now we have the average rating for every restaurant!

Pivot tables

- Used for data restructuring, aggregation, general analysis
- Grouped Aggregation
  - Average Rating Per Restaurant [YelpReviewsSample]
    > Select all data > Data > Pivot Table..
    > Rows > Add Field > name
    > Values > Add Field > stars > summarize by: AVERAGE
  - Number of Reviews Per Restaurant [YelpReviewsSample]
    > Values > Add Field > stars > summarize by: COUNT

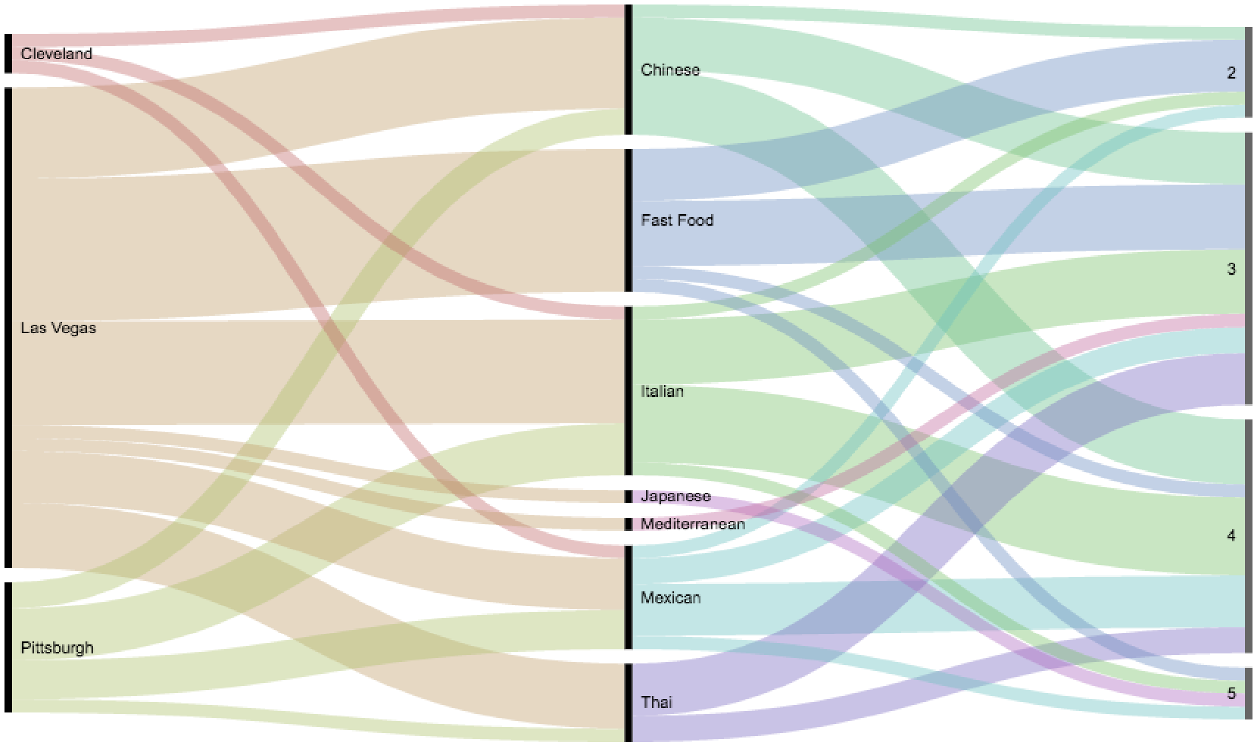
Data Visualization - Basic

Note: Visualizations (charts) in Google Sheets have a way to go to catch up with Excel.

Bar charts
Useful when one axis is categories and the other is numeric
Pie charts
Useful when comparing sizes of categories
Scatterplots
Useful when both axes are numeric (or at least ordered)

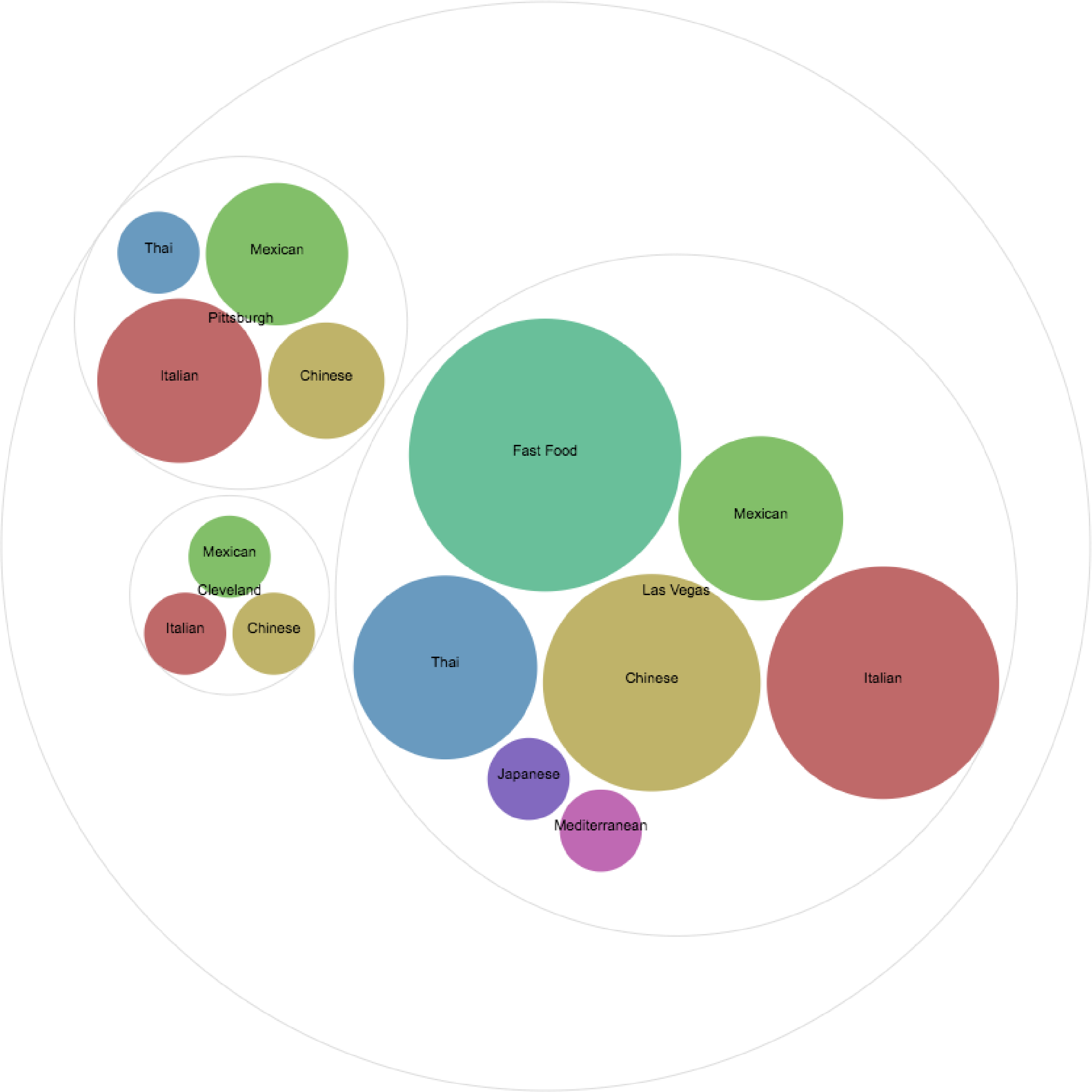
More advanced/exotic visualizations using Raw tool (http://raw.densitydesign.org/)
> First copy paste data / upload .csv

- Alluvial diagram
  - Drag "city" into STEPS
  - Drag "category" into STEPS
  - Drag "Average Rating" into STEPS



- Circle packing

○ Drag “city” into HIERARCHY
○ Drag “category” into HIERARCHY
○ Drag “category” into Color


Thai
Mexican
Pittsburgh
Italian
Chinese
Mexican
Cleveland
Italian
Chinese
Fast Food
Mexican
Las Vegas
Thai
Chinese
Italian
Japanese
Mediterranean