MLDS 411
DATA VISUALIZATION

Winter 2024 Labs

Alice Zhao

Northwestern | McCORMICK SCHOOL OF ENGINEERING

Schedule

Week	Date	Time	Tableau Topics	Python Topics
1	January 12	11am	Tableau Review	Library Overview
2	January 19	12pm	Joining, Blending	Plotly, Dash
3	January 26	11am	Groups, Hierarchies, Sets	Bokeh
4	February 2	11am	Parameters, Tool Tips	Wordcloud
5	February 9	11am	Dashboards, Stories & Actions	NetworkX
6	February 16	11am	Maps	GeoPandas, Folium
7	February 23	11am	Calculated Fields, Calculations	Plotnine
8	March 1	11am	Forecasting, Clustering	Python + Tableau

Lab 2 Follow Ups

- Can you use both joins and relationships in the same workbook?
  - **Short answer**: Yes
  - **Long answer**:
    - **Joins** are good for combining simple data sets into a single table
    - **Relationships** are good for matching one-to-many rows to keep tables separate

- Tableau doesn't seem to have as many features or as much flexibility as some other business intelligence tools
  - **Short answer**: Correct
  - **Long answer**:
    - Remember that each language / tool has its strengths and weaknesses
    - **Python**: you can do so many things with Python, but the learning curve is steep
    - **Tableau**: this was specifically built for a non-technical audience, so it will be limited

Today's Theme: Organizing Data

Hierarchies

- Organize data into a hierarchical structure
- Allows for drill down analysis

Going from top to bottom, or less detailed to more detailed

Groups

- Categorize data into groups
- Allows you to create higher-level categories

Going from bottom to top, or more detailed to less detailed

Sets

- Create a subset of data based on a condition
- Allows you to compare the subset of data to the rest of the data

Comparing a set of data points to all other data points

Demo: Hierarchies

Why create a hierarchy?

- Allows you to quickly drill down into levels for analysis

Common applications of hierarchies

- Built in hierarchy: Dates (discrete vs continuous)
- See the drill down: Categories (category >> subcategory)
- Avoid aggregations: Geographical data (country >> state >> city)

How can you create a hierarchy?

- Drag one dimension on top of another within the Data tab
- Select multiple dimensions >> Hierarchy >> Create Hierarchy

Demo: Groups

Why create a group?

- Allows you to bucket together multiple categories into fewer categories

Common applications of groups

- Group many small, miscellaneous categories into a group called “Other”
- Color multiple categories the same color on a visualization

How can you create a group?

- Select categorical values within a visualization >> Group
- Select a dimension or measure on the Data tab >> Create >> Group

Demo: Sets

Why create a set?

- Allows you to compare a subset of data with the rest of the data

Common applications of sets

- Color multiple data points the same color on a visualization
- Create custom filters (show top x performing items, etc.)

How can you create a set?

- Fixed sets: select specific data points on a visualization >> Create Set
- Dynamic sets: select data points based on a condition >> Create >> Set
  - Can create a Combined Set with a Calculated Field (Create >> Calculated Field)

Comparison: Groups vs Sets

Groups	Sets
Often used to create **custom categories**	Often used to create **custom filters**
Used to create **higher-level** categories for dimensions	Used to create **subsets** to compare with the rest of the data
Only works with **one dimension** at a time	Can work with **multiple dimensions** at a time
Groups are **fixed**	Sets can be **dynamic**
Members can only belong to **one group**	Members can belong to **multiple sets**

Exercise: Organizing Superstore Data

- Create a heatmap that shows the average profit each quarter for all states, and add on a hierarchy that contains region and state
- Compare the monthly profit trends between a group of the south and west region versus a group of the central and east region
- Create a scatter plot of profit versus sales for all sub-categories
  - Highlight the sub-categories where the profits are negative using a fixed set
  - Highlight the sub-categories where the quantity is more than 1000 using a dynamic set

Organizing Data Summary

1. If you'd like to create a hierarchical structure in your data

- Create a hierarchy

2. If you'd like to create custom, high-level categories

- Create some groups

3. If you'd like to compare a subset of data versus the rest

- Create a subset

Data Visualization in Python Overview

Standard Libraries

Mapping

Interactive Visualizations

Specialty

amueller / word_cloud

Bokeh

Bokeh

- Plotting library that allows you to create interactive plots
- Example bokeh charts: https://docs.bokeh.org/en/latest/docs/gallery.html

Plotly vs Bokeh

- **Plotly**: more popular, bigger team, easy to get up and running
- **Bokeh**: more niche, smaller team, can handle more complex data

Demo

- Candy_Analysis_Plotly_vs_Bokeh.ipynb