CIS545 Recitation 3:

Integration and Visualization

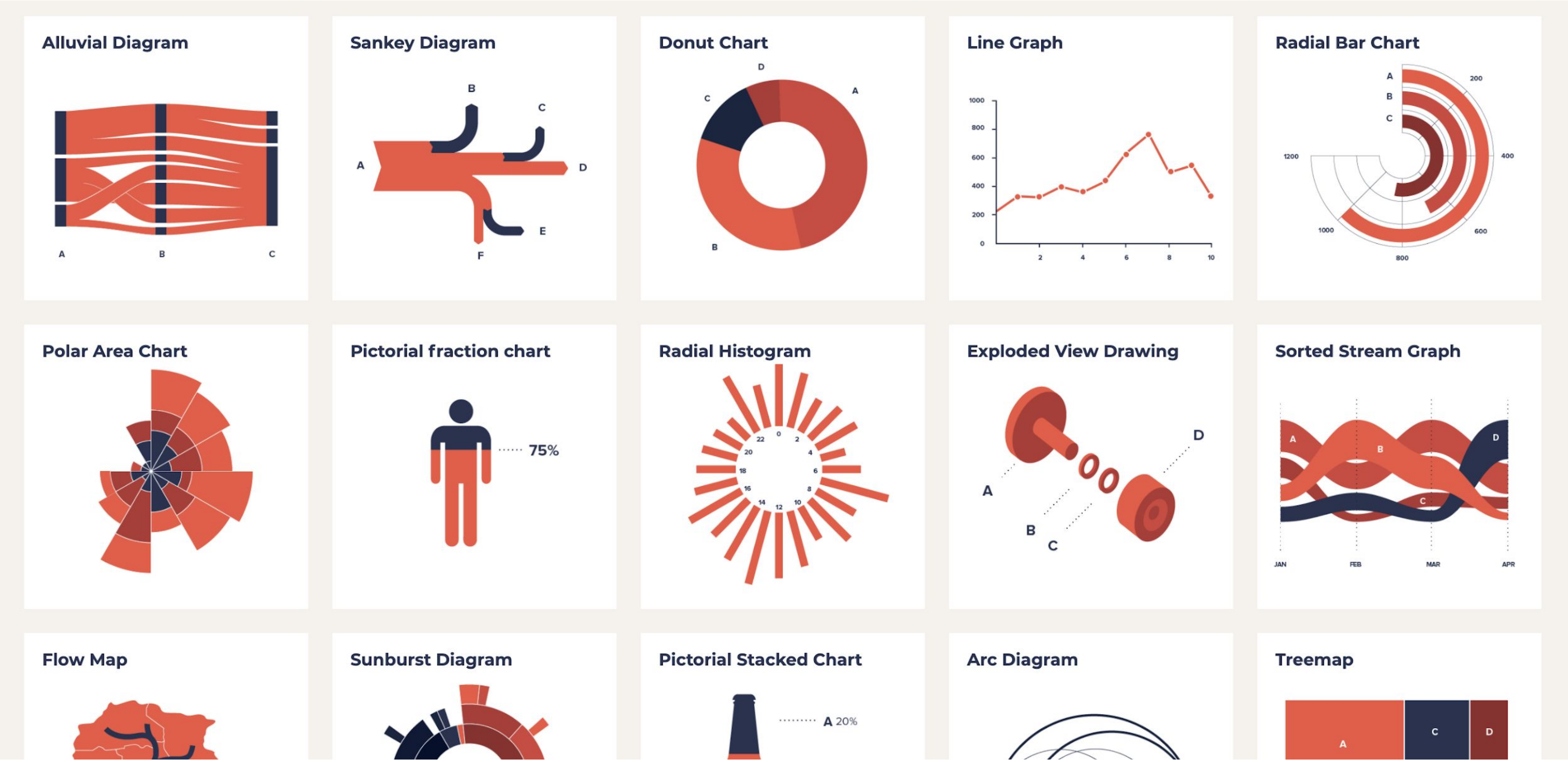
15 September 2023

Data Visualization

https://informationisbeautiful.net/

General Steps for Data Viz

1. Look at your data
   a. What variables do we have?
   b. How many entries?
2. Identify the message and its components
   a. What question am I trying to answer?
   b. Who am I trying to answer it for?
   c. What variables do I need?
3. Select your chart
   a. Identify the type of relationship (comparison, distribution, etc.)
   b. https://datavizproject.com/
4. Refine
   a. Add labels, titles, legends, etc.
   b. Pick meaningful colors



Alluvial Diagram

Sankey Diagram

Donut Chart

Line Graph

Radial Bar Chart

Polar Area Chart

Pictorial fraction chart

Radial Histogram

Exploded View Drawing

Sorted Stream Graph

Flow Map

Sunburst Diagram

Pictorial Stacked Chart

Arc Diagram

Treemap

Python Tools for Data Visualization

An Overview of:

- pandas
- seaborn
- ggplot
- plot.ly

Non-Python Tools:

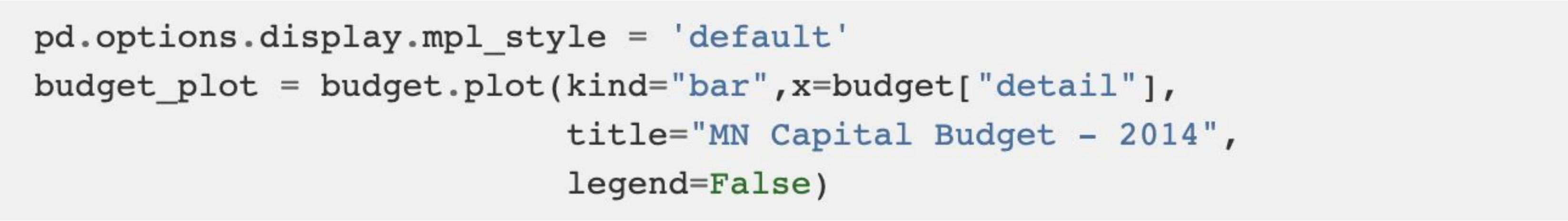
- Tableau
- D3.js

How to Use Data Visualization Tools

- **Don't:** learn each tool from scratch!
    - Instead, always go from purpose/idea → tool
- **Do: find a template for your visualization and substitute in your variables**
    - think of data visualization packages not as scripting languages, but as formatting tools
    - each tool has its own way of breaking down the components of the graph, so it's a lot easier to base it off a template than starting from scratch
    - **WARNING:** the code blocks for visualization will initially look VERY involved **but don't be discouraged and slowly break it down into its functional components**
    -
-

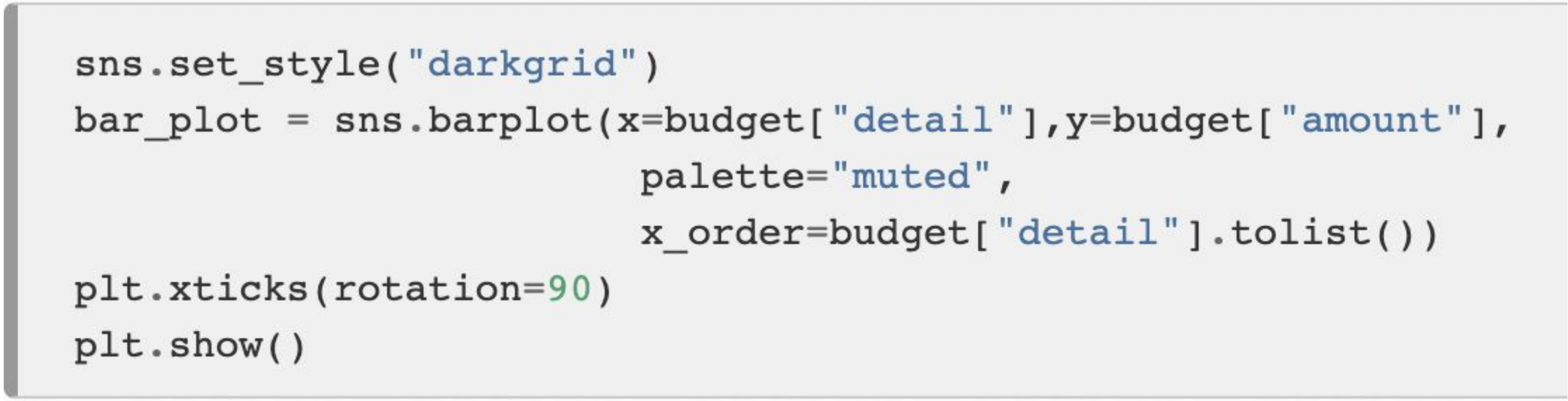
pandas and seaborn

- Pandas: built-in visualization component for in the library
  - User Guide



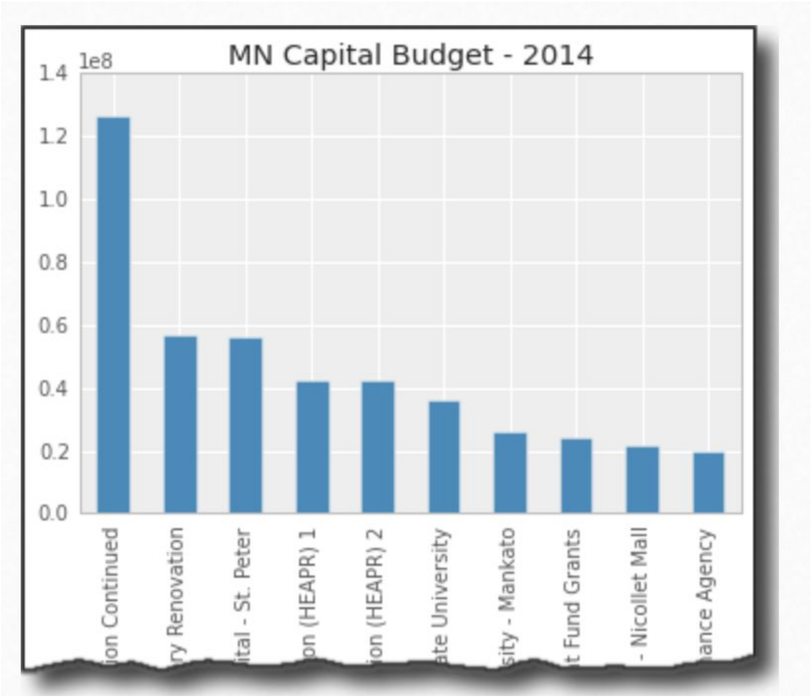
```
pd.options.display.mpl_style = 'default'
budget_plot = budget.plot(kind="bar",x=budget["detail"],
                          title="MN Capital Budget - 2014",
                          legend=False)
```

- Seaborn: aims to make default data visualizations more visually appealing with more formatting options
  - User Guide

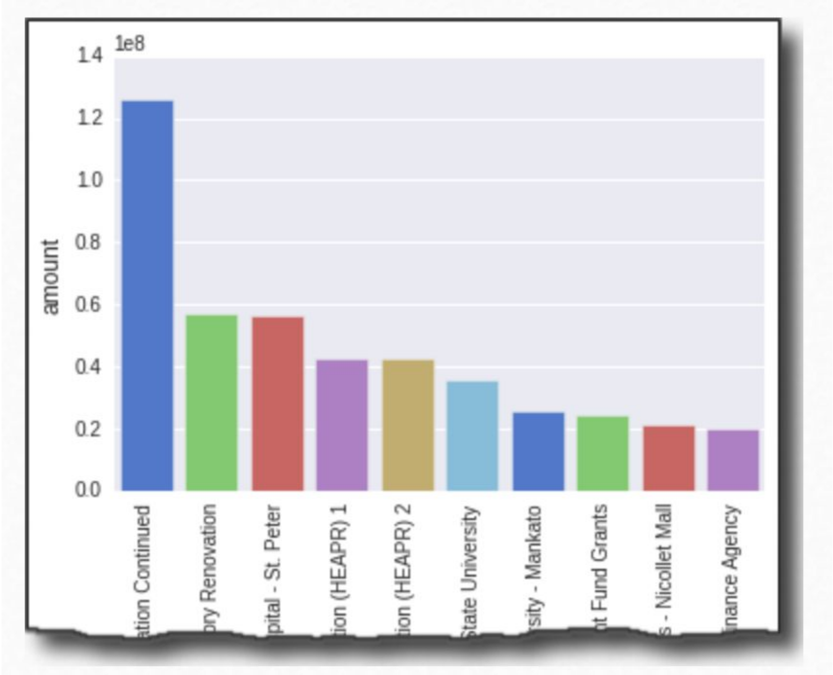


```
sns.set_style("darkgrid")
bar_plot = sns.barplot(x=budget["detail"],y=budget["amount"],
                       palette="muted",
                       x_order=budget["detail"].tolist())
plt.xticks(rotation=90)
plt.show()
```

pandas

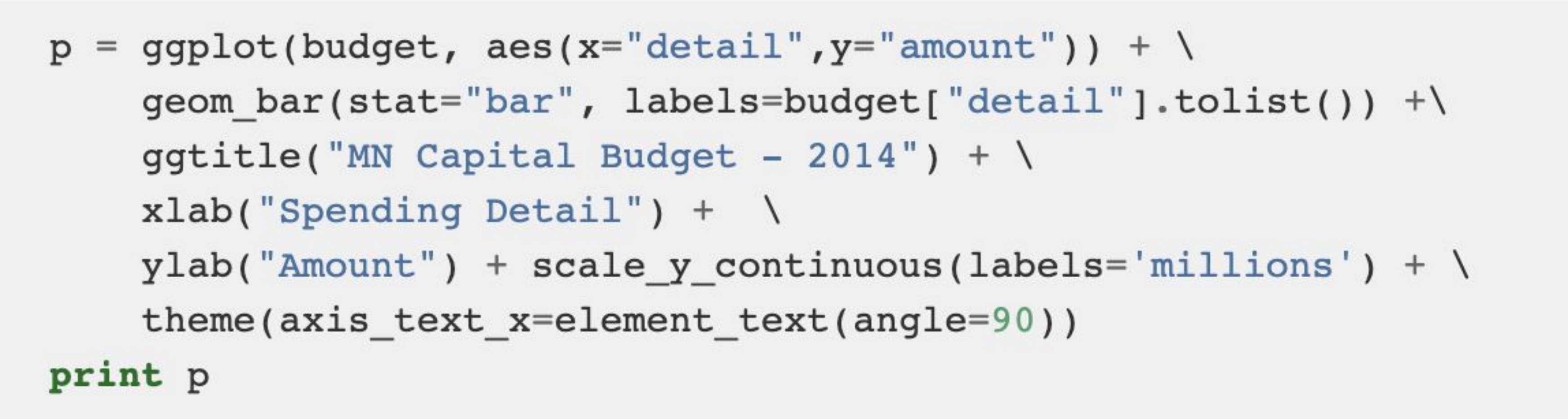


seaborn



ggplot

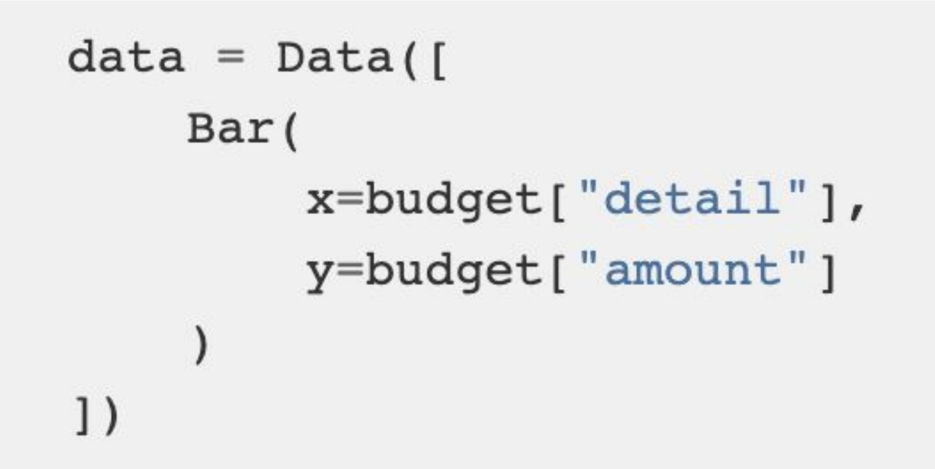
- Originally created for R
- Aims to break up graphs into semantic components such as scales and layers
- User Guide



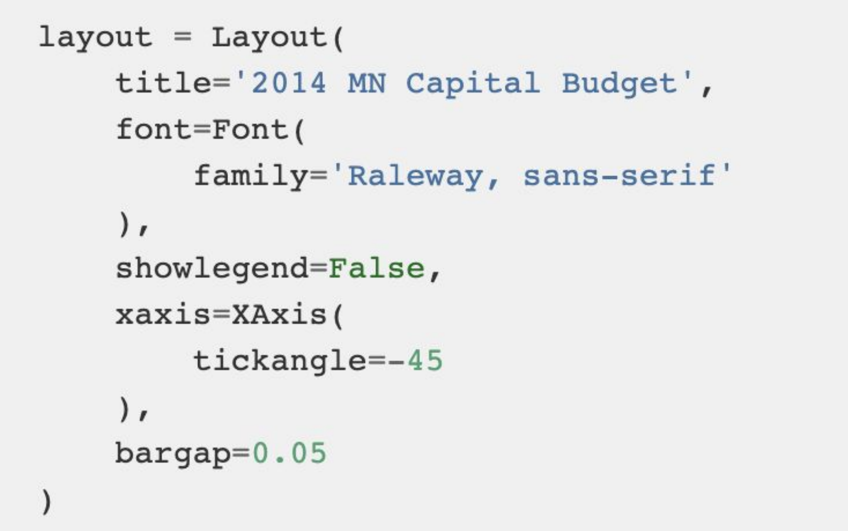
```
p = ggplot(budget, aes(x="detail",y="amount")) + \
    geom_bar(stat="bar", labels=budget["detail"].tolist()) +\
    ggtitle("MN Capital Budget - 2014") + \
    xlab("Spending Detail") +  \
    ylab("Amount") + scale_y_continuous(labels='millions') + \
    theme(axis_text_x=element_text(angle=90))
print p
```

plot.ly

- Built on top of D3.js, a JavaScript library for producing interactive data visualizations in web browsers
- Available for both Python and R
- Components of graphs are treated like the concept of an “object” common in languages in like Java
- 3D Scatter Plots Example



```
data = Data([
    Bar(
        x=budget["detail"],
        y=budget["amount"]
    )
])
```



```
layout = Layout(
    title='2014 MN Capital Budget',
    font=Font(
        family='Raleway, sans-serif'
    ),
    showlegend=False,
    xaxis=XAxis(
        tickangle=-45
    ),
    bargap=0.05
)
```

Data Viz Demo

(Back to the Notebook!)

Thank You!