Nolan Rittenberg

Data Science Workflow: Data collection → Data Prep → Data Viz → Data Analysis → Data storytelling

Univariate Viz:

- Layer = geometric elements (ex. lines, Points)
- Theme = Plot aesthetics
- ggplot() is used for data viz

Bivariate Viz:

- Response Variable = Variable we want to explain
- Predictors = Variables that might explain something else (ex. elevation of a hike)

MultiVariate Viz:

- Use of facets, Color, Shape can add more variables
- Add layers and change theme.

Spatial Viz

- Point Maps: Plot location of indiv. Obs.
- Contour Maps: Plot density of distrib. of Obs.
- Choropleth Maps: Plot outcomes in diff. regions

Effective Viz

- No "One" right Viz
- Viz can be objectively wrong, bad, or ugly
- Professionalism, Accessibility, Ethics, details.

Wrangling and dates

- arrange(): arrange the rows according to some column
- filter(): filter out or obtain a subset of the rows
- Select(): Select a subset of Columns
- Mutate(): Mutate or Create a Column
- Summarize(): Calculate the numerical Summary of a Column
- group_by(): group the rows by a specific column
- |>: Pipe ("and, then")
- ==: equal to
- !=: not equal to
- > greater than, >= greater than or equal to
- < less than, <= less than or equal to
- %in% c(***, ***): a list of multiple values

How to use geom_:

Univariate Categorical = _bar
Univariate Numerical = _hist, _density, _boxplot
Bivariat 2 Categorical = _bar(Position = "dodge")
Bivariate Cat. and Num. = _boxPlot, _violin
Bivariat 2 Num. = _Point, _smooth
Multi 2 Cat. 1 num = _bar + fill/aes()
Multi 1 Cat 2 num = _boxPlot + facet_wrap(), _point(aes(color=cat))
Multi 3 Num = _Point (aes(color, size = variable_3)

Reshaping:

- Pivot_Wider: increases # of Columns
- Pivot_longer: increases # of rows

Joining: Join 2 datasets into 1

Mutate Joins:

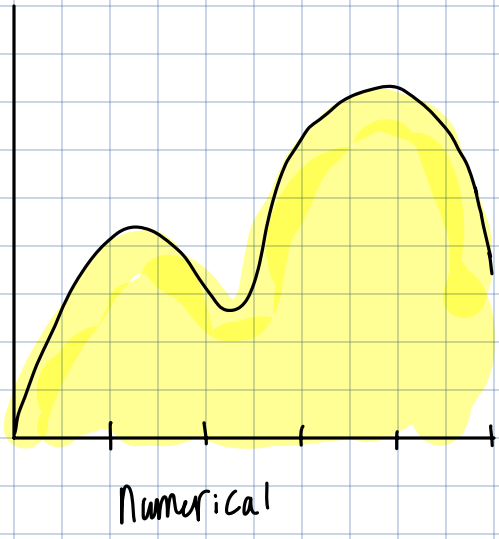
- left_join()
- inner_join()
- full_join()

Filtering Joins

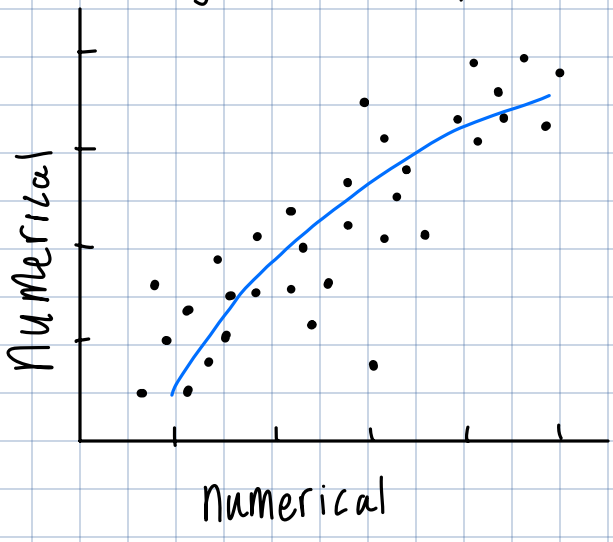
- semi_join()
- anti_join()

Visualizations

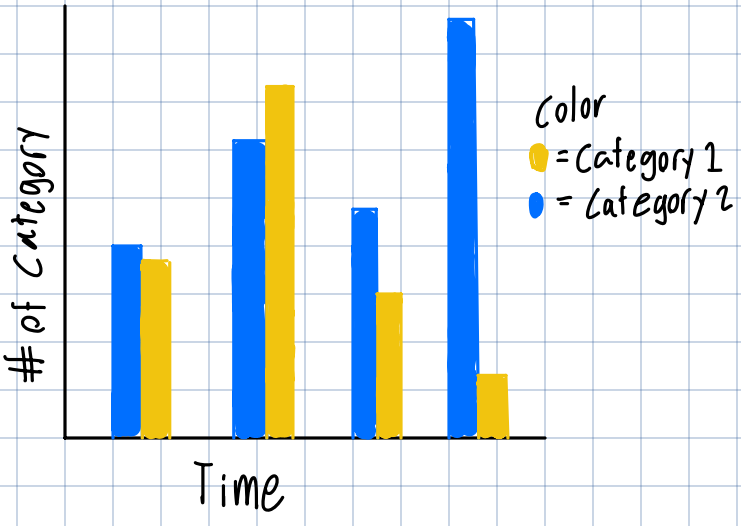
geom_density



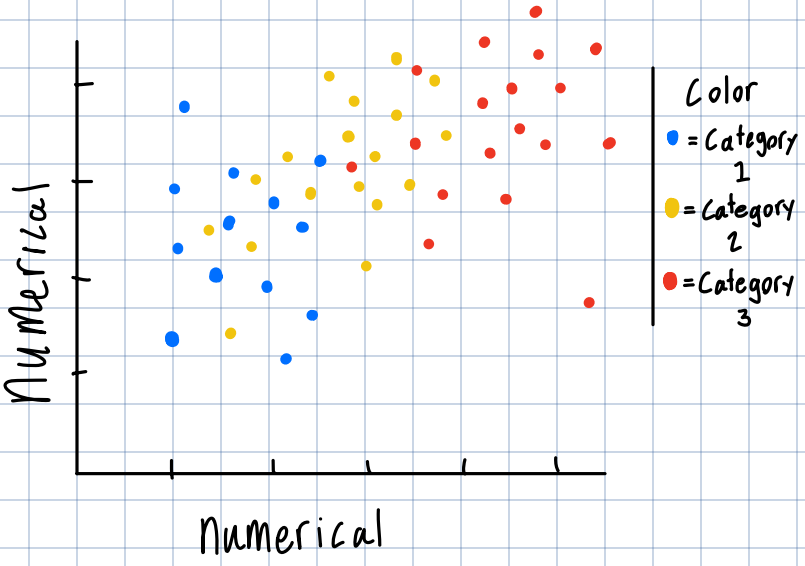
geom_Point() + geom_Smooth()



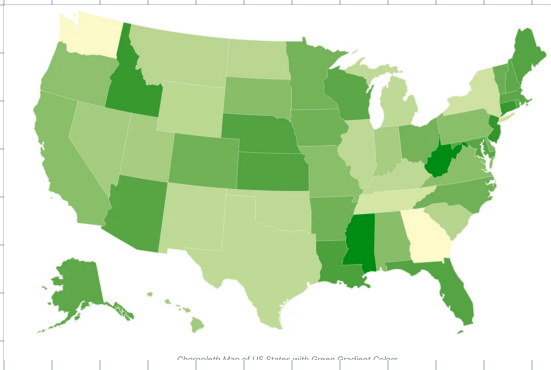
geom_bar(Position="dodge")



geom_Point()



Choropleth Map



darkness shows density of an observation in a state

Point Map: Each Point represents an observation