Determining When and Why to Use Multivariate Analysis

Guillermo Fernández
DATA SCIENTIST

@guillermo_ai

Summary

- Get used to multivariate analysis techniques
- Understand when and why to use them
- Understand the insight we can get from each technique
- Perform multivariate analysis techniques with Python

Finding Relationships in Data

Multivariate Analysis Goal

Express relations by a statistical dependence

When possible

Characterize dependence by a numerical coefficient

Independent of the units of measure of the variables

Covariance

Measure of joint probability of 2 random variables

Shows tendency in a linear relationship

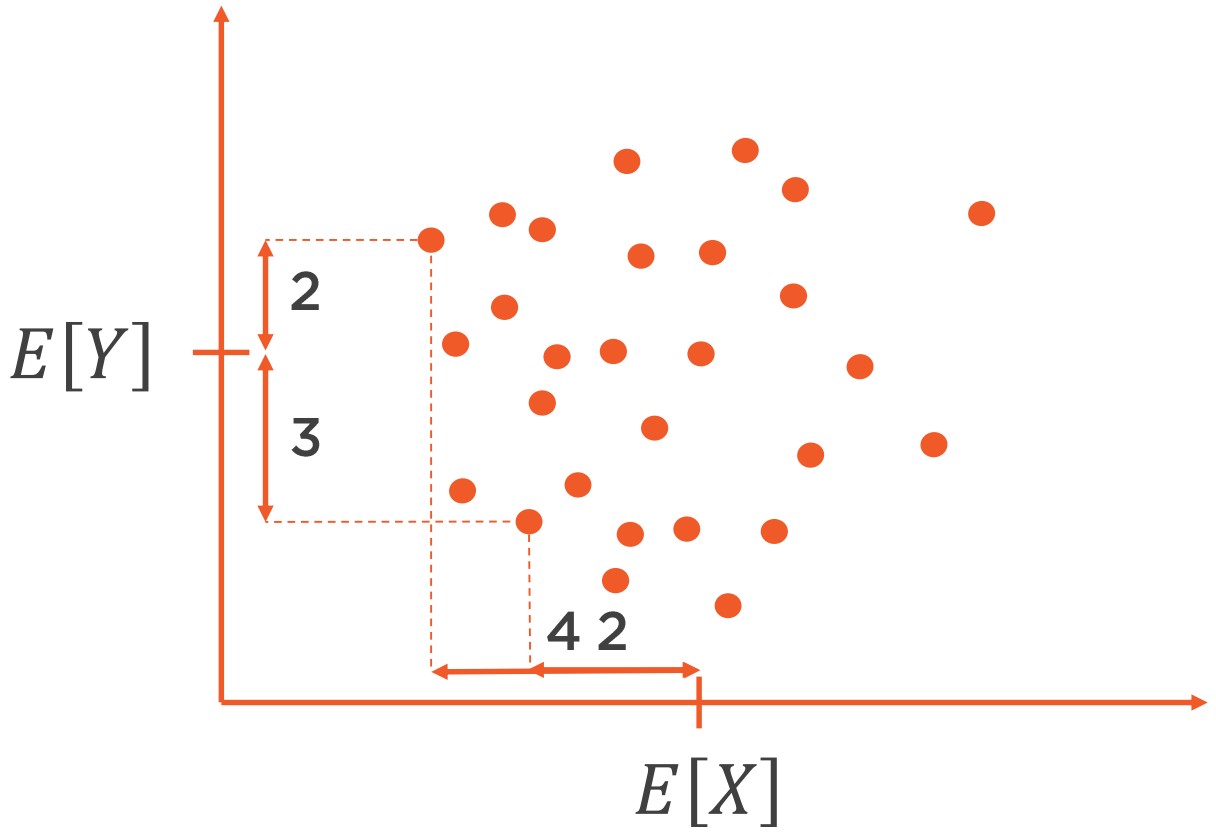
Positive or Negative

Has units

Not easy to interpret strength of relationship

Covariance Intuition

$$\sigma(X,Y) = \ cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$



$$\frac{4*2 + 2*3 + \cdots}{\text{number of points}}$$

Correlation

Useful because it can indicate a predictive relationship

Tells the strength and direction of relationship

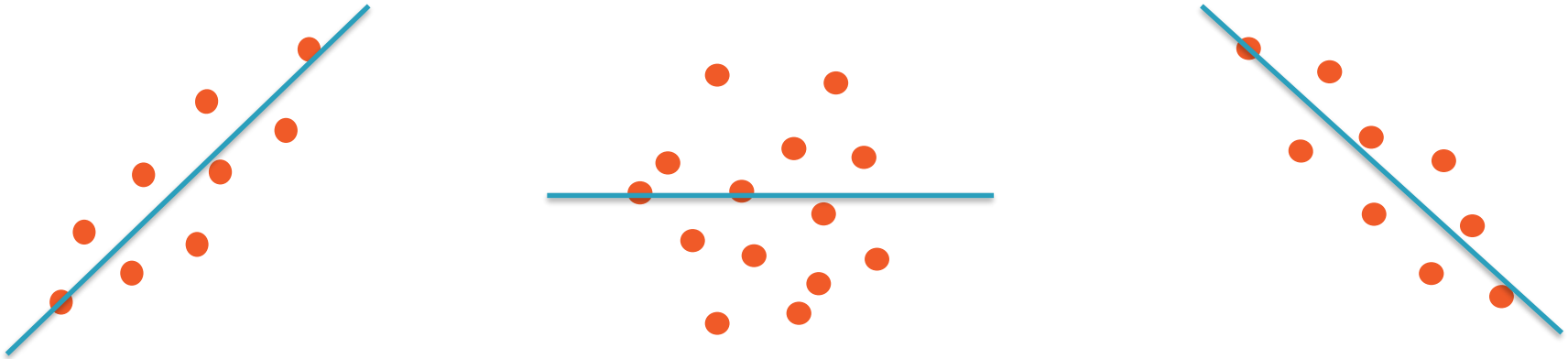
Positive or Negative

Commonly refers to a linear relationship

Pearson Correlation

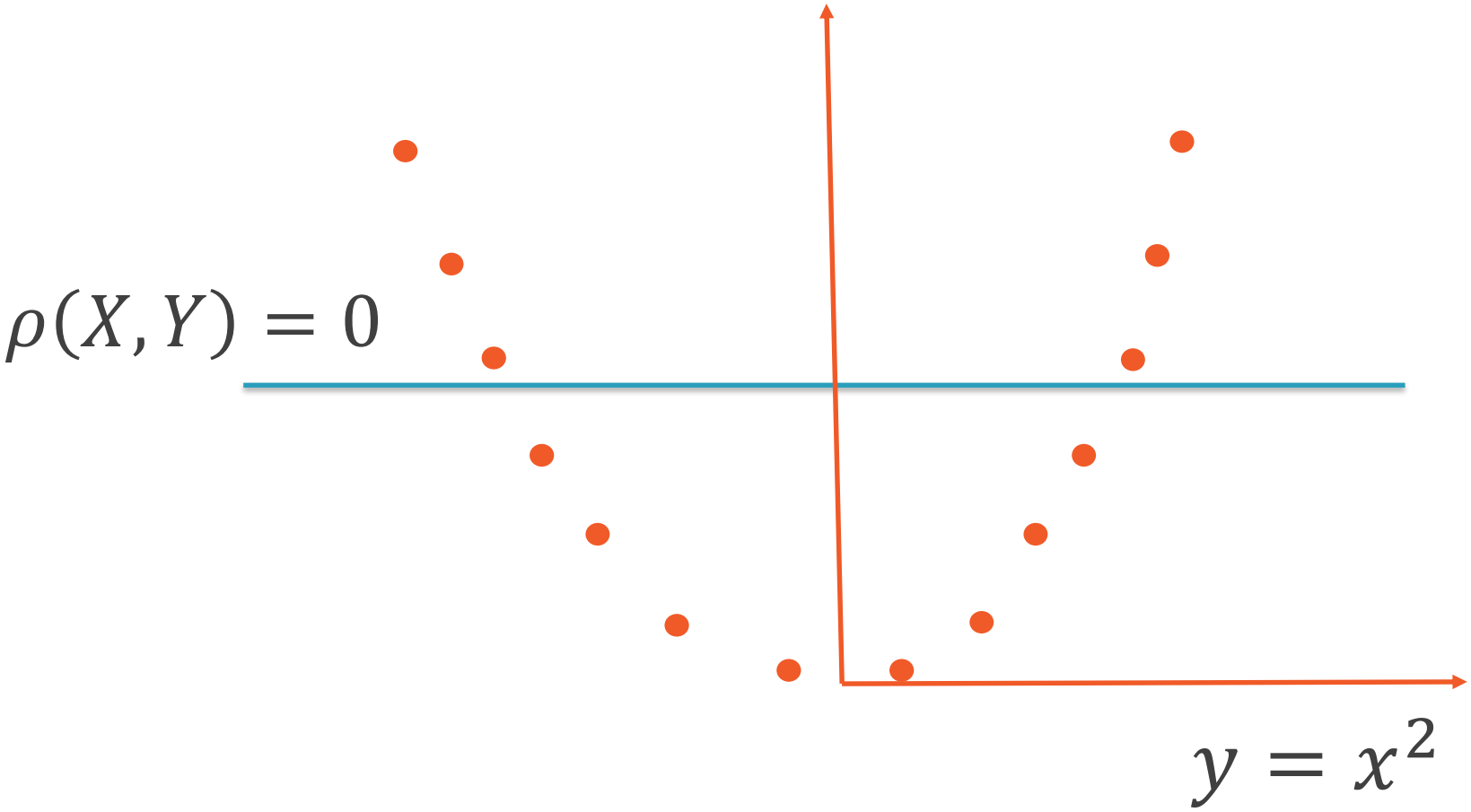
Correlation

$$\rho(X,Y) = \ corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

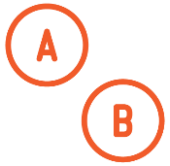


1 0 -1

Correlation



Advise About Correlation



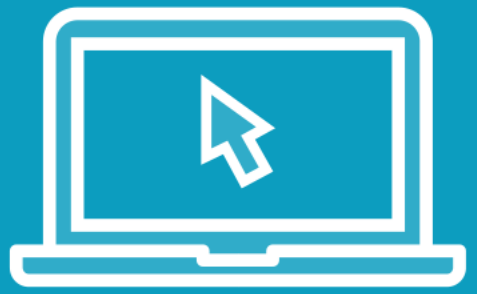
If 2 variables are independent, their correlation is 0

A correlation of 0 does not imply that 2 variables are independent

Cannot replace visual examination of the data

The presence of correlation is not enough to infer causation

Demo



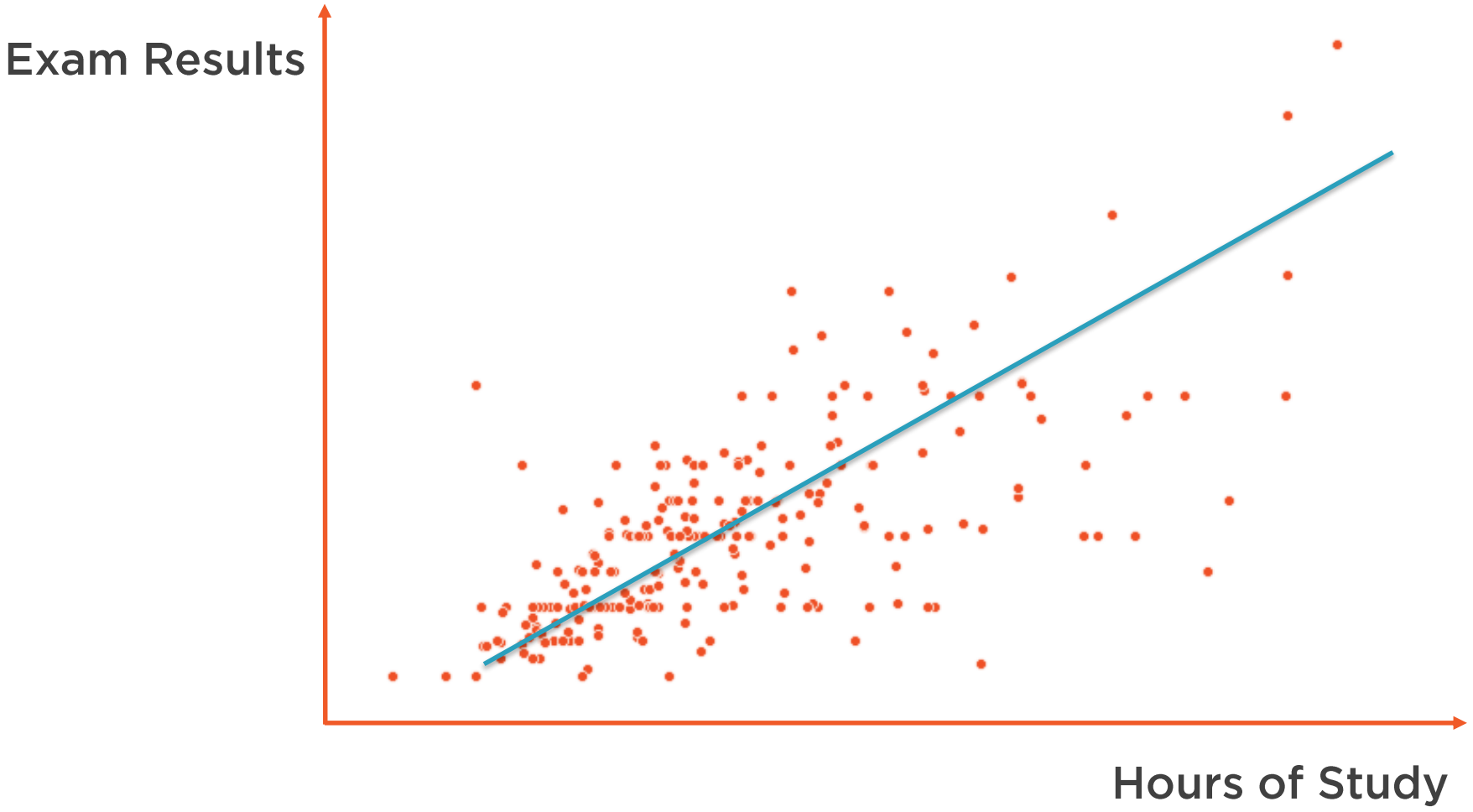
Learn how to compute covariance and correlation with Python

Using Python packages

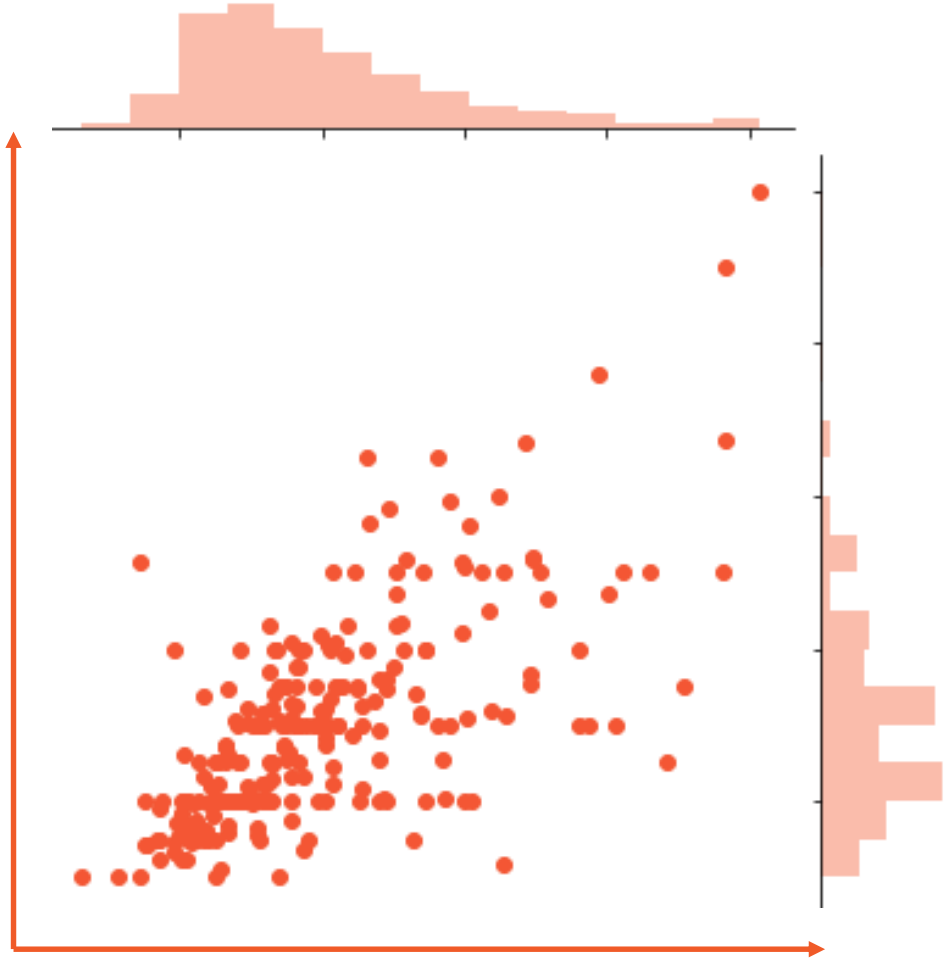
- Numpy
- Pandas

Multivariate Distribution Plots

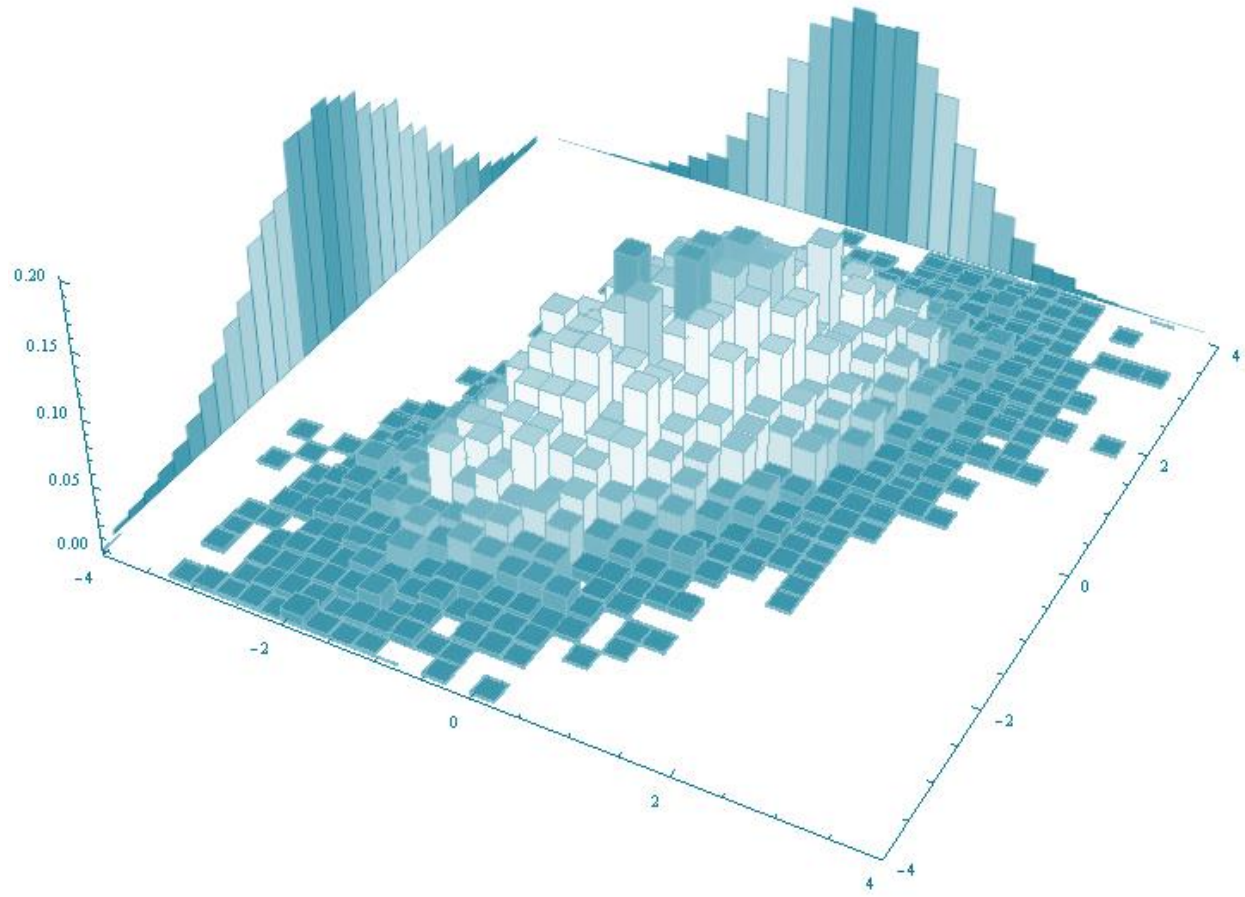
Scatter Plot



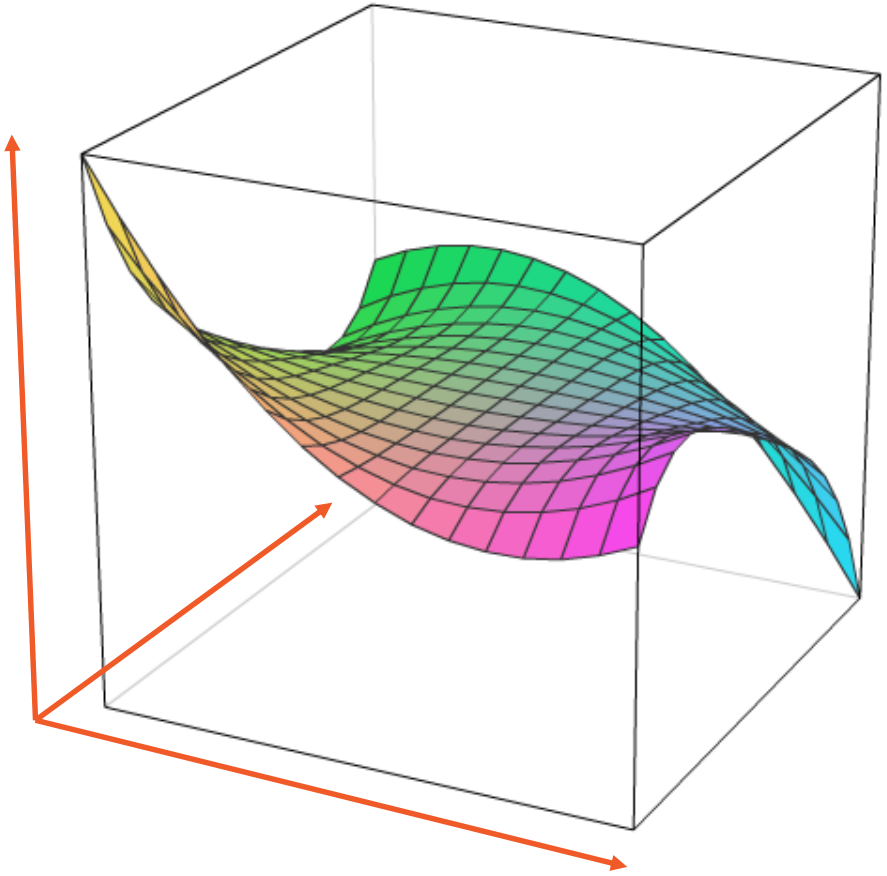
Joint Distribution Plot



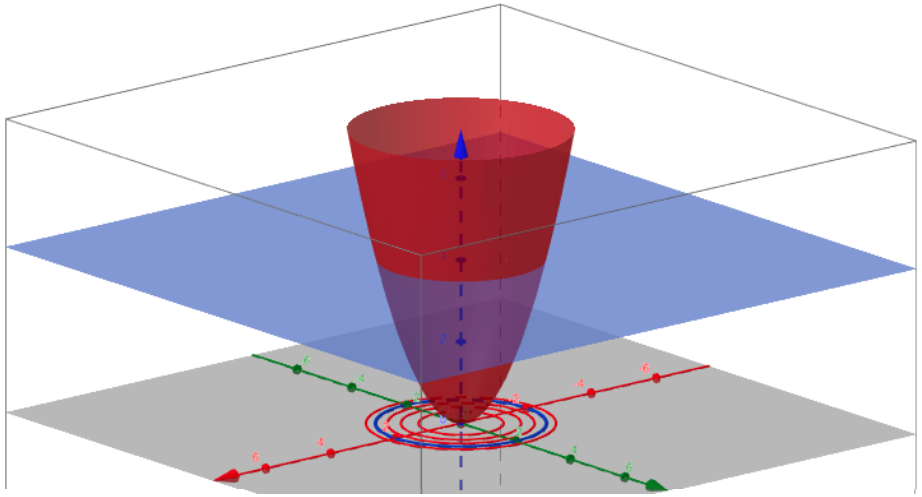
Stereogram or Bivariate Histogram

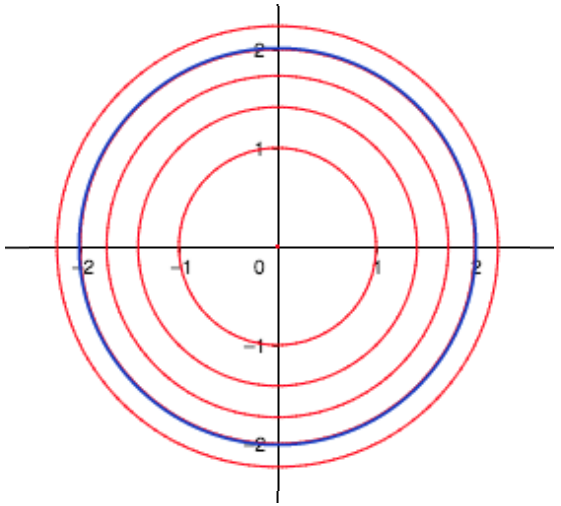


Surface Chart

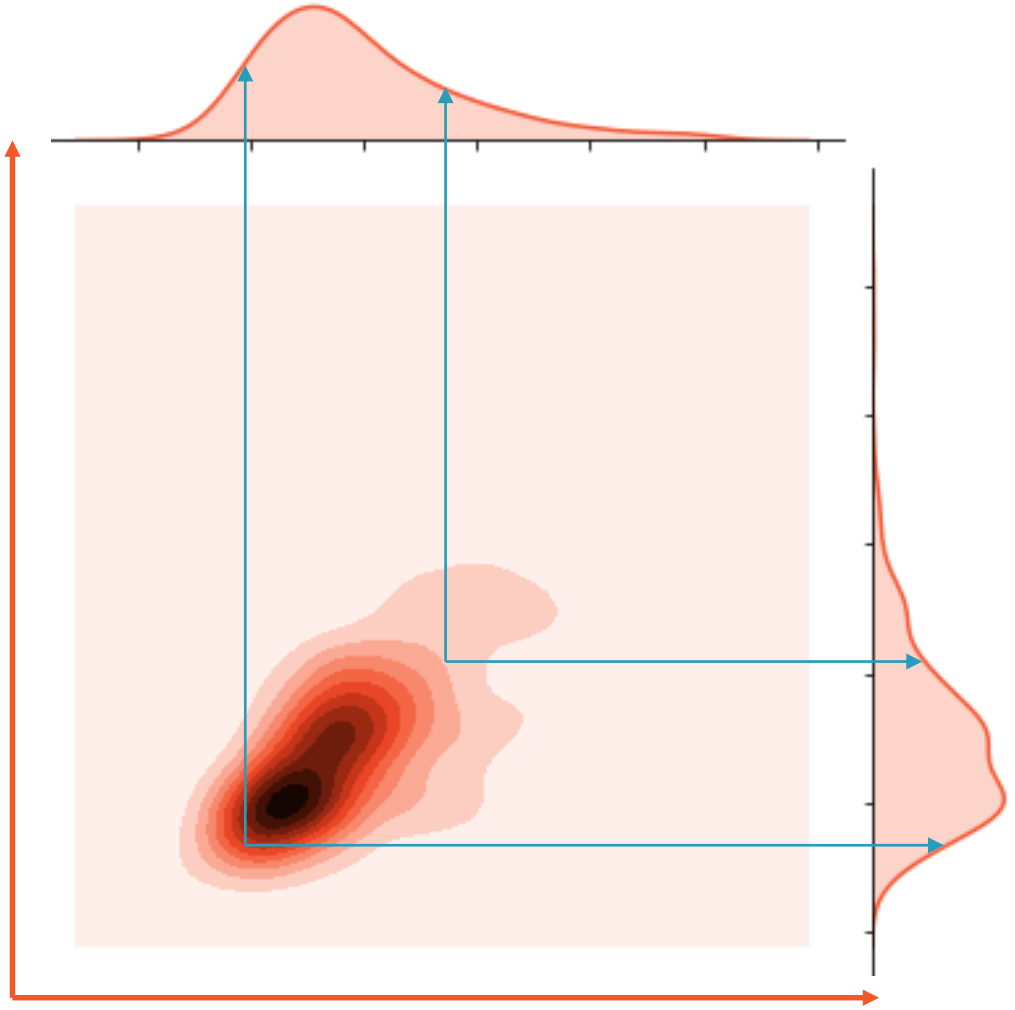


Level Curves

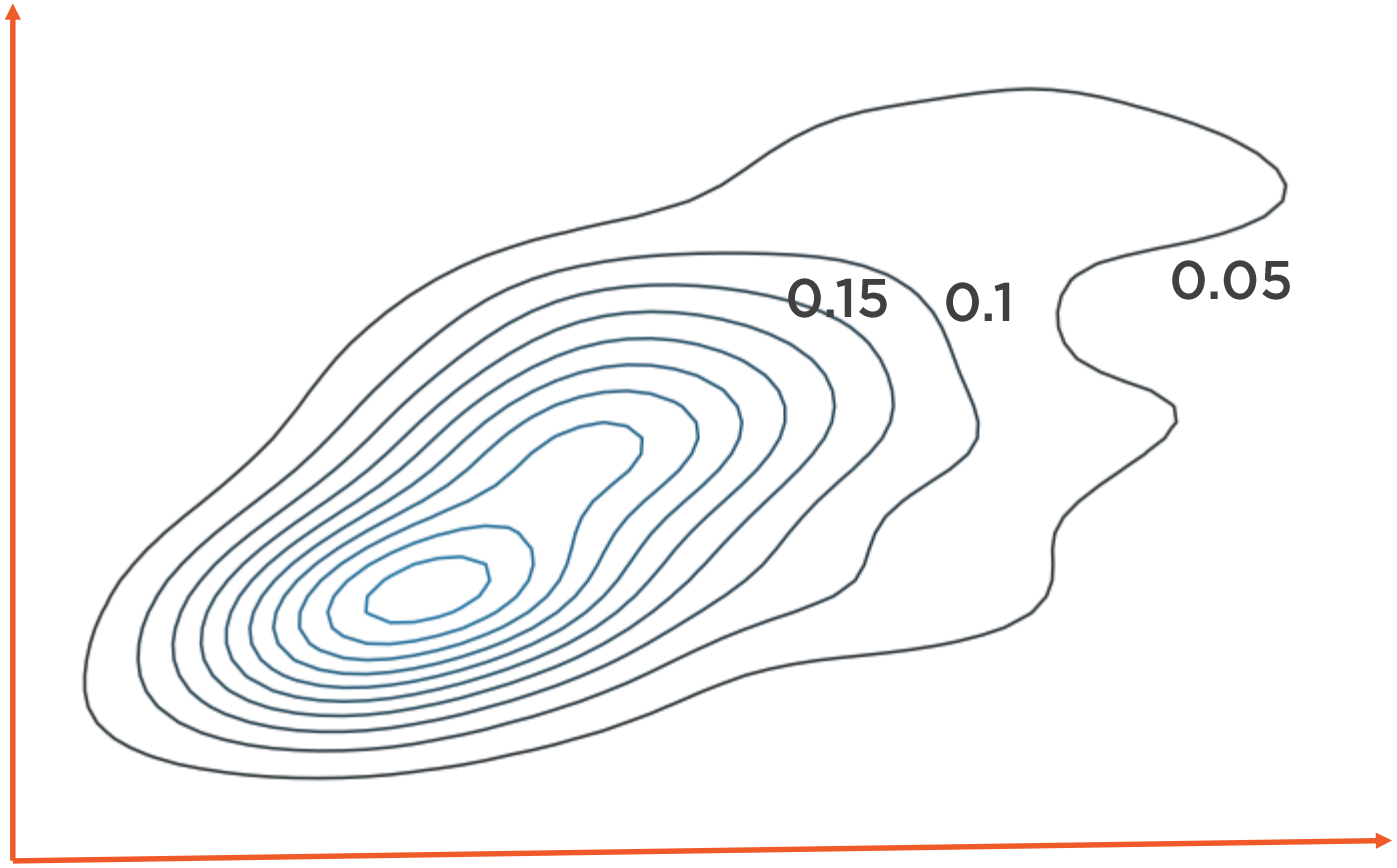




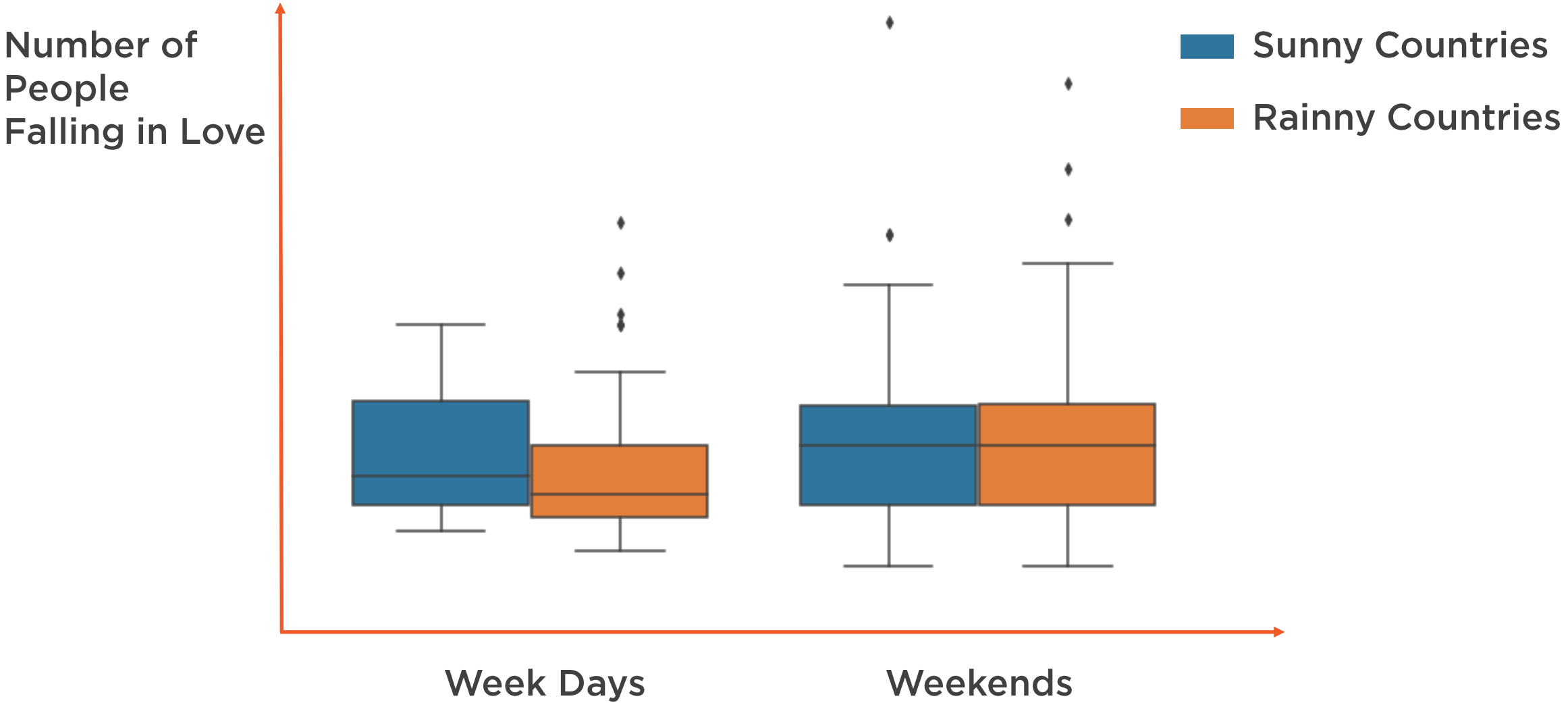
Level Curves



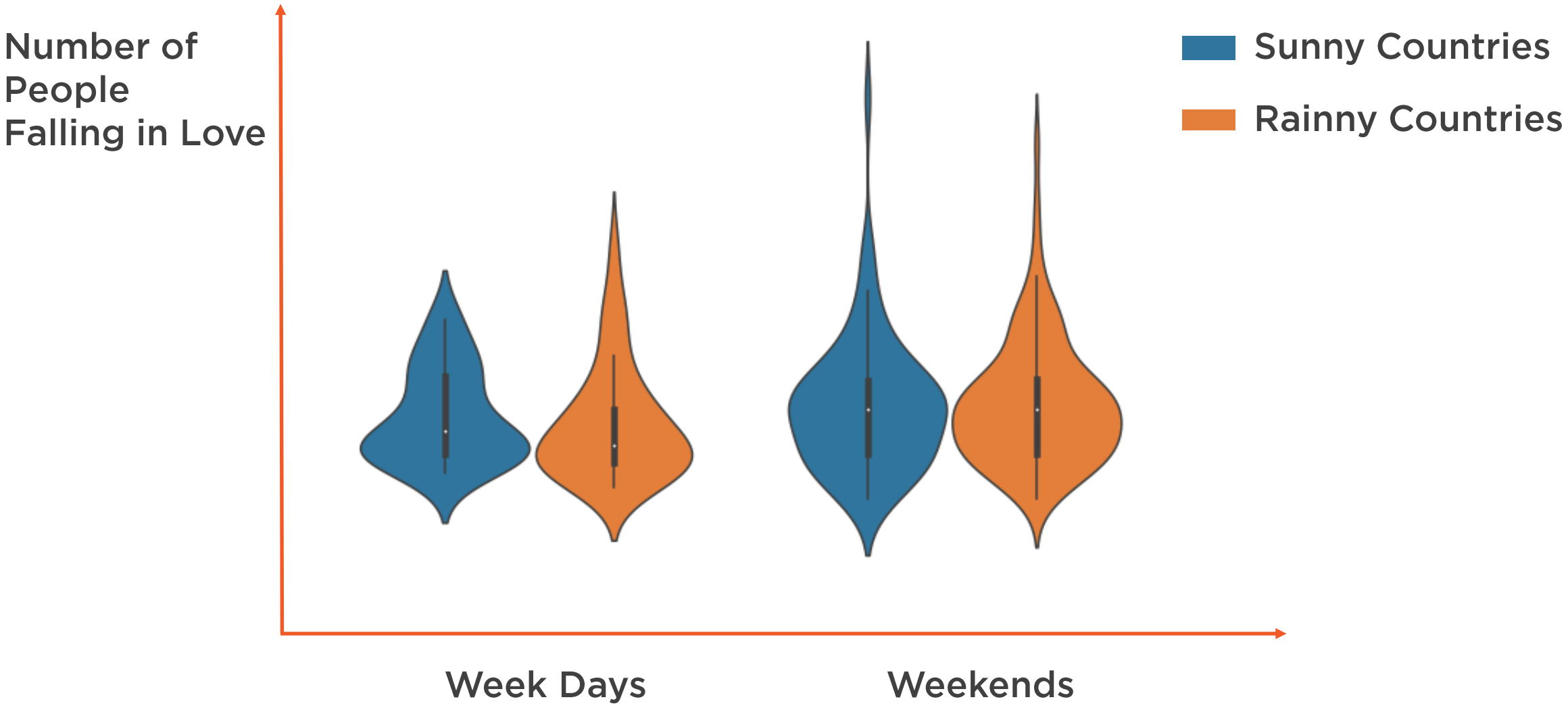
Level Curves



Box Plots


Number of People Falling in Love
Sunny Countries
Rainny Countries
Week Days
Weekends

Violin Plots



Demo

Learn how to plot multivariate comparison charts with Python

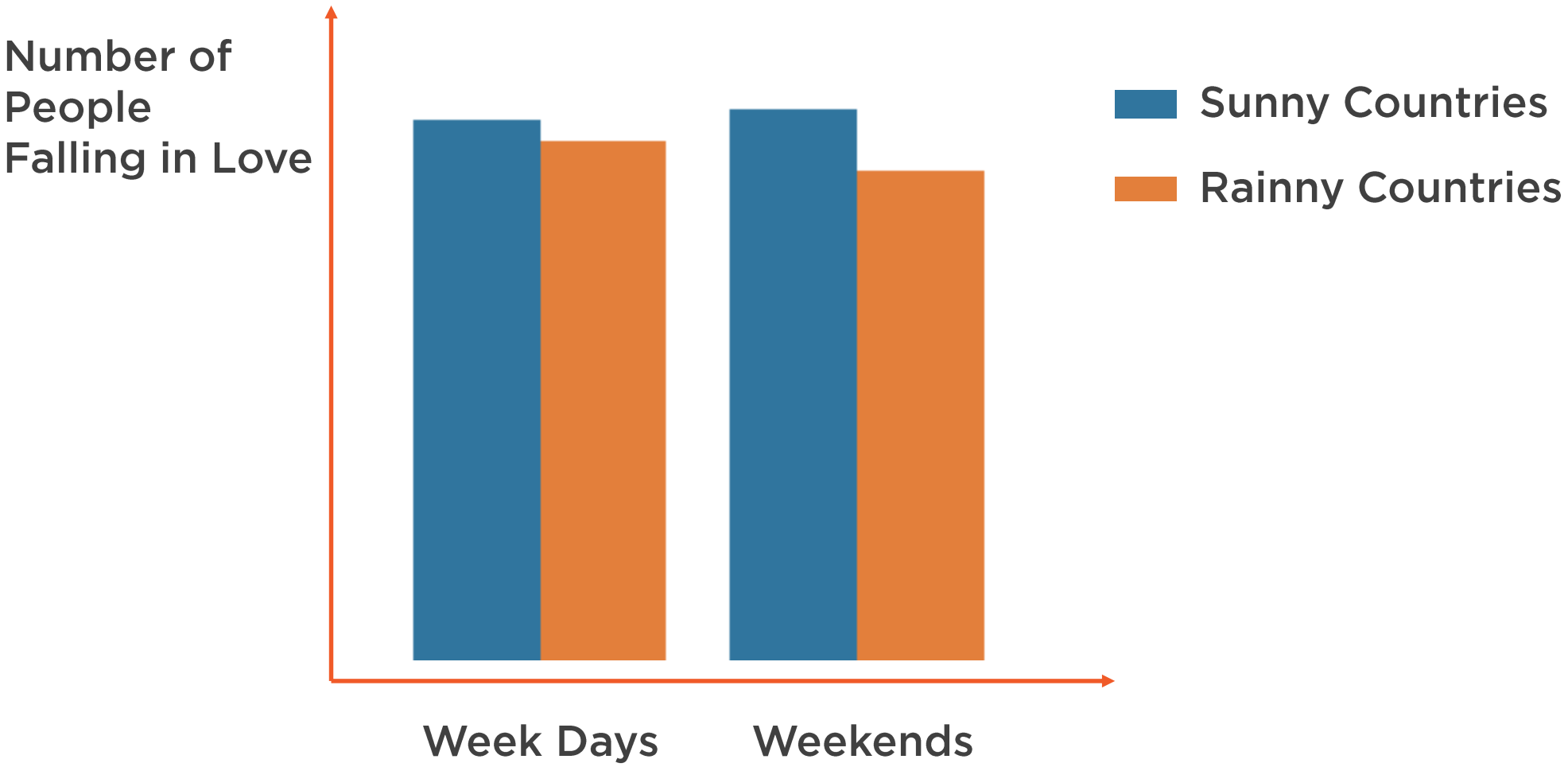
Using Python packages

- Matplotlib
- Seaborn

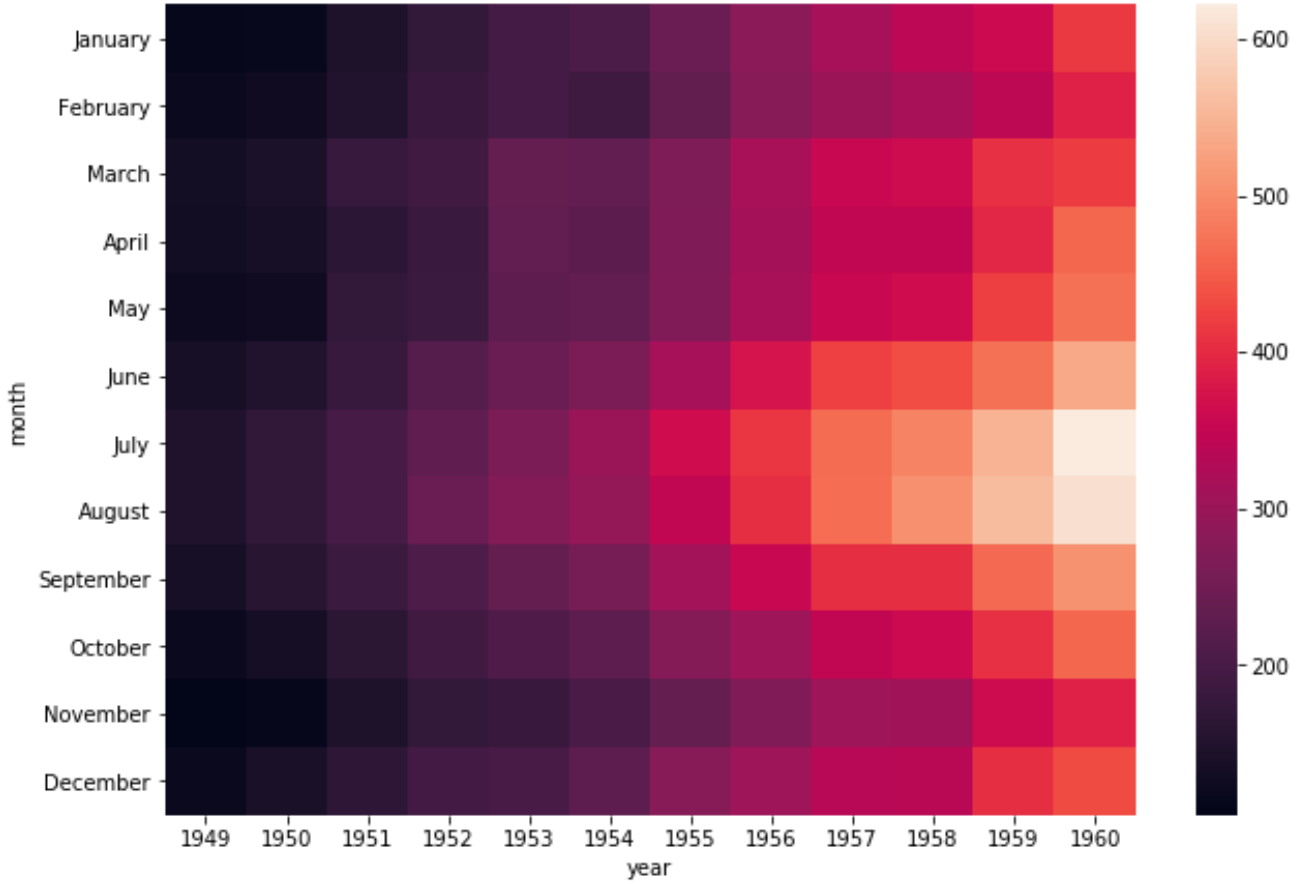
Learn how to customize some simple graph aesthetics

Multivariate Comparison Plots

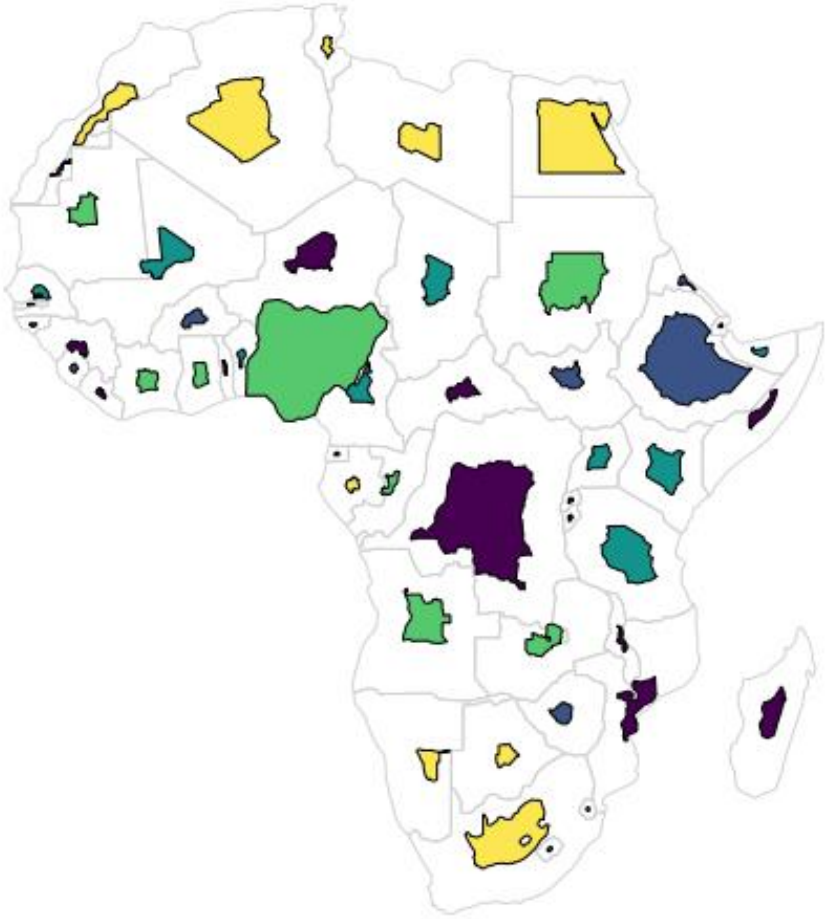
Bar Chart


Number of People Falling in Love
Sunny Countries
Rainny Countries
Week Days
Weekends

Heatmap


January
February
March
April
May
June
July
August
September
October
November
December
month
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
year
600
500
400
300
200

Cartograms



Demo

Learn how to plot multivariate comparison charts with Python

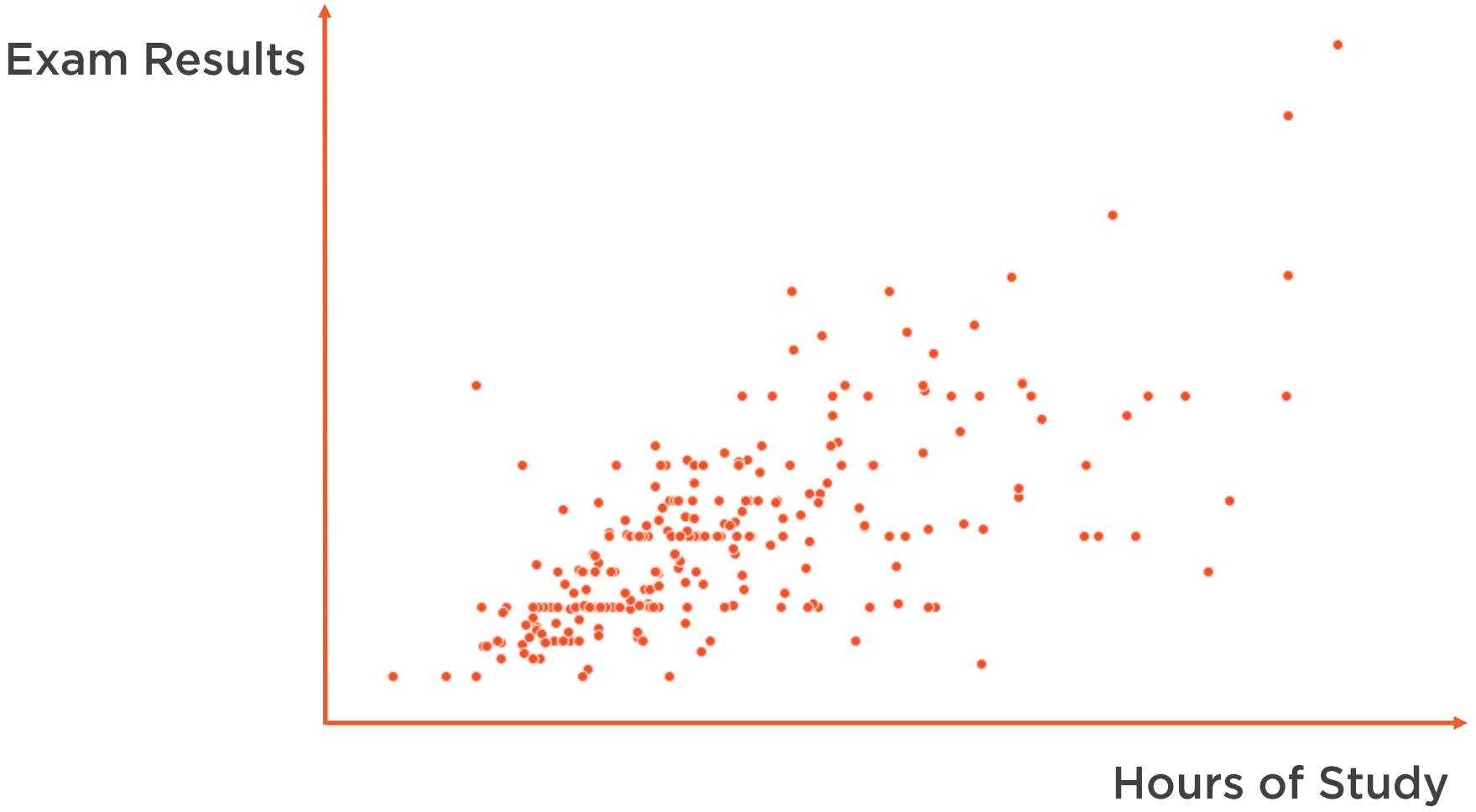
Using Python packages

- Matplotlib
- Seaborn
- Pandas
- Geopandas
- Geoplot

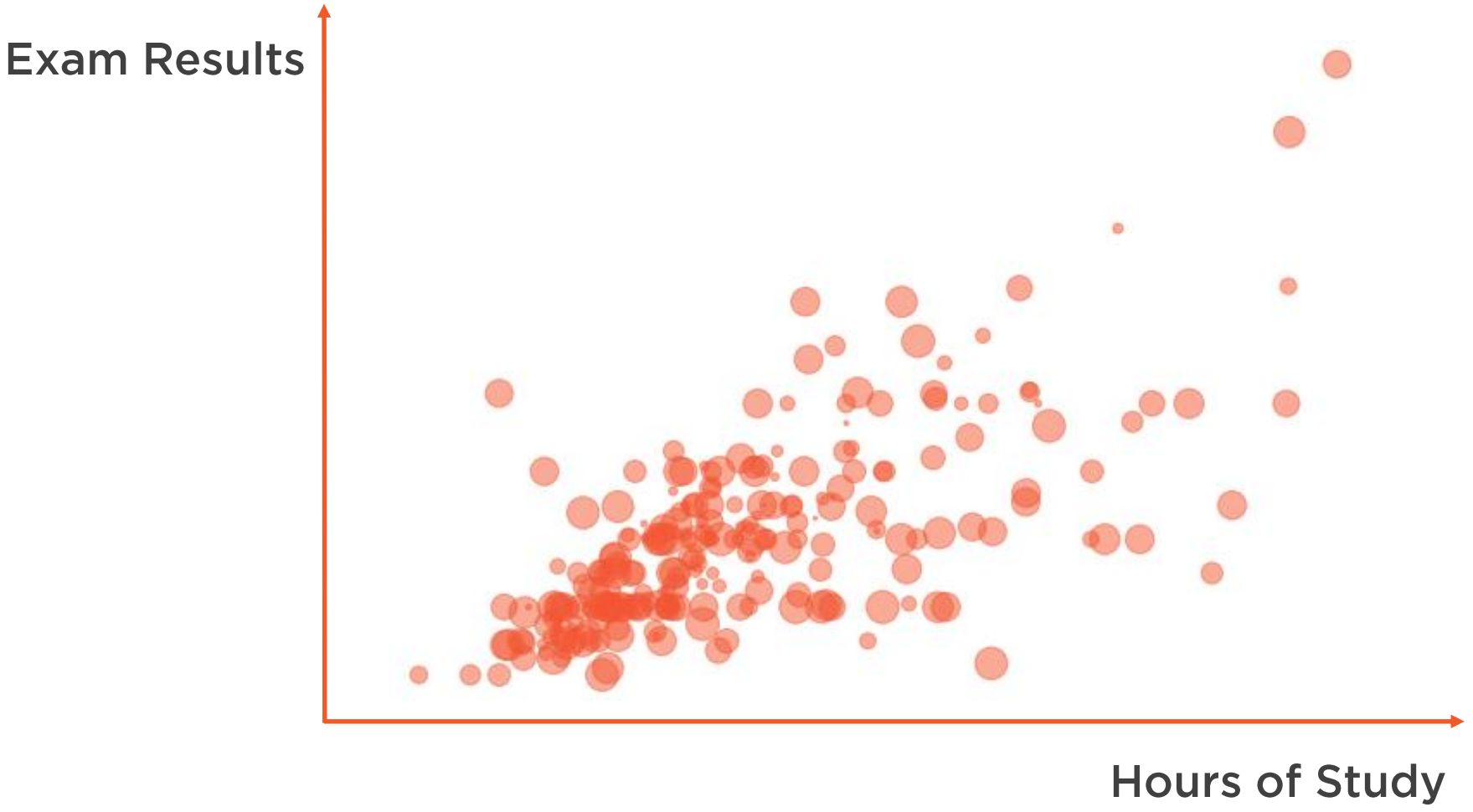
Learn how to customize some simple graph aesthetics

Multivariate Relationship Plots

Scatter Plot



Bubble Plot



Demo

Learn how to plot multivariate comparison charts with Python

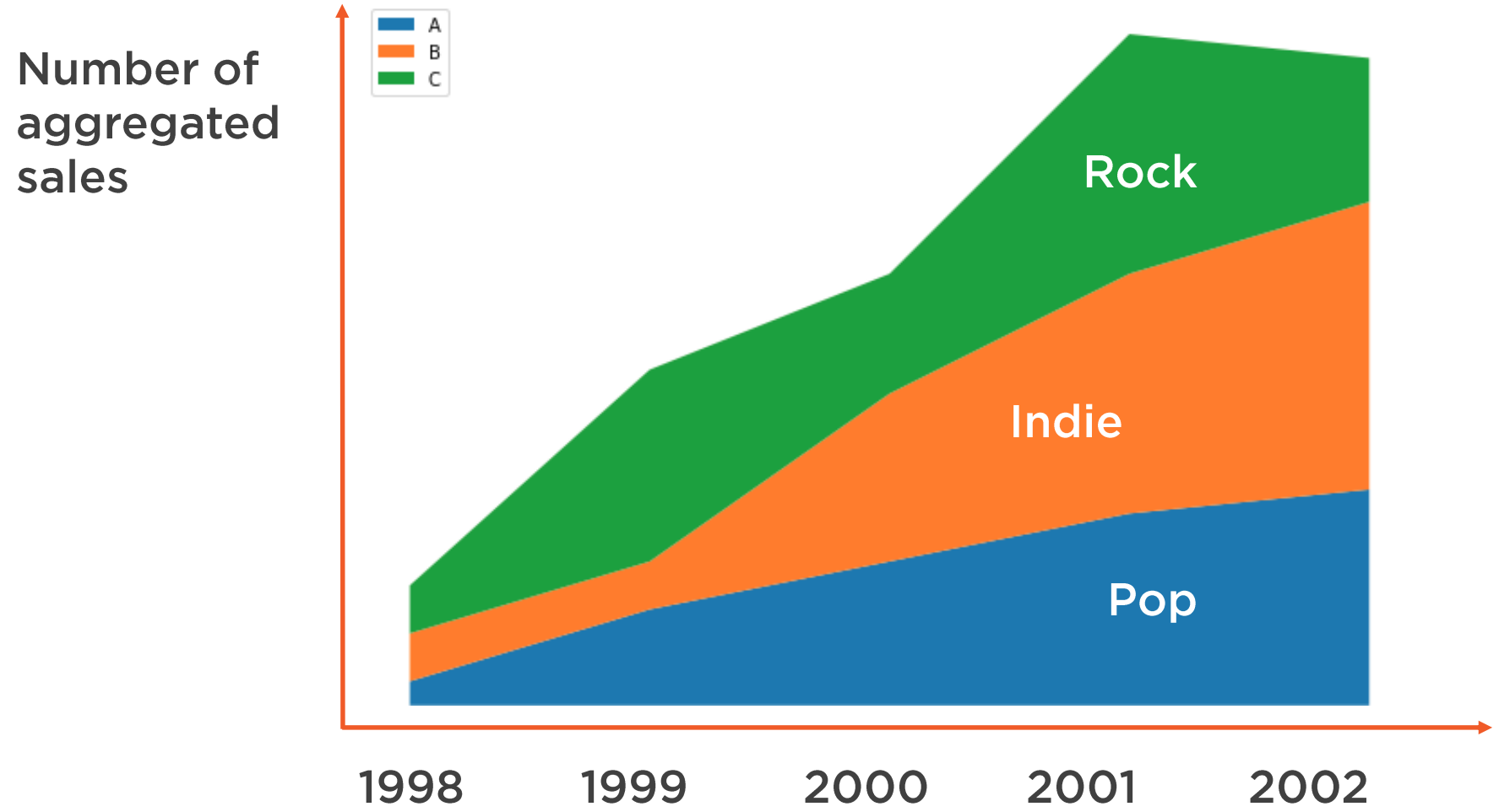
Using Python packages

- Matplotlib
- Seaborn

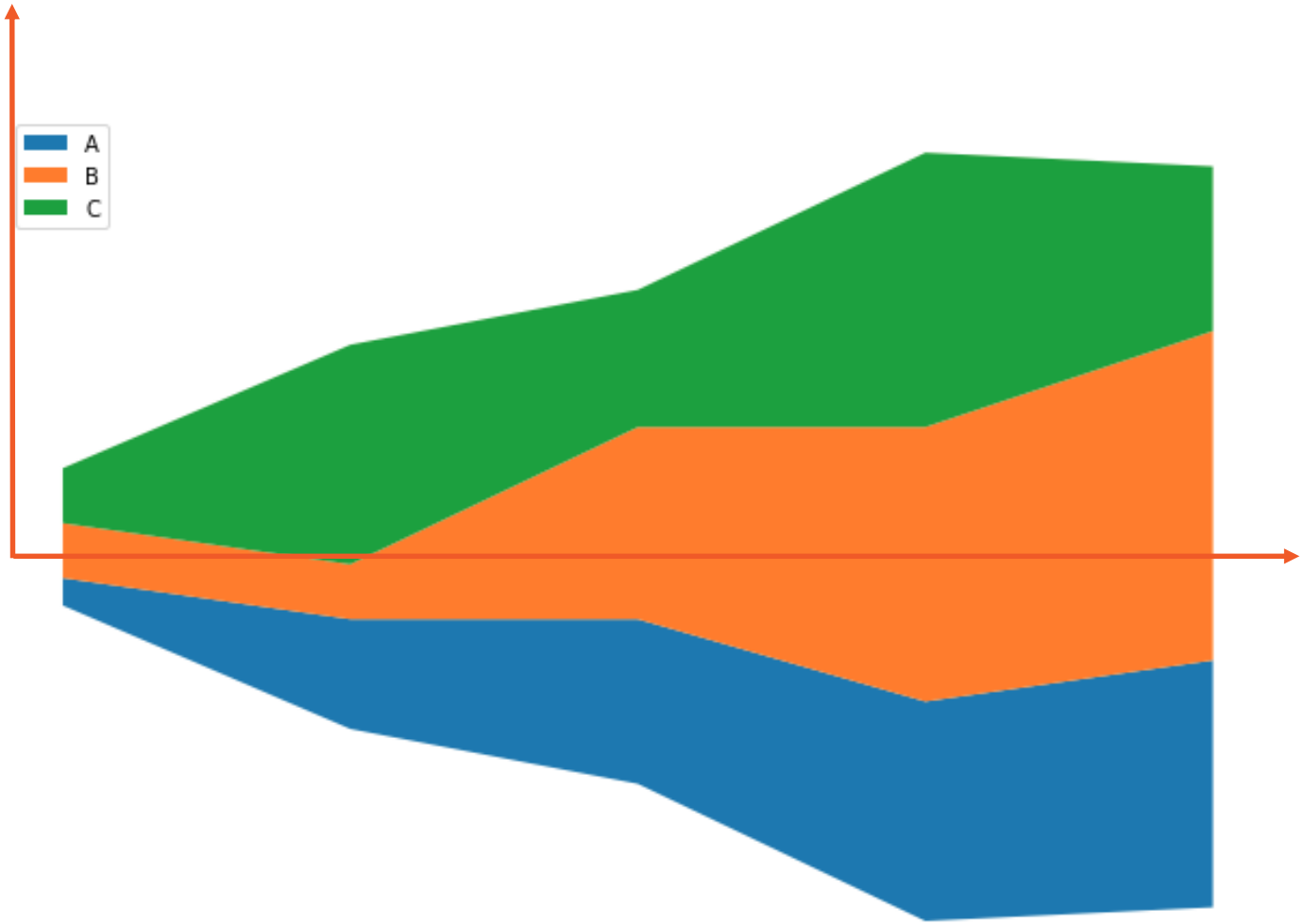
Learn how to customize some simple graph aesthetics

Multivariate Composition Plots

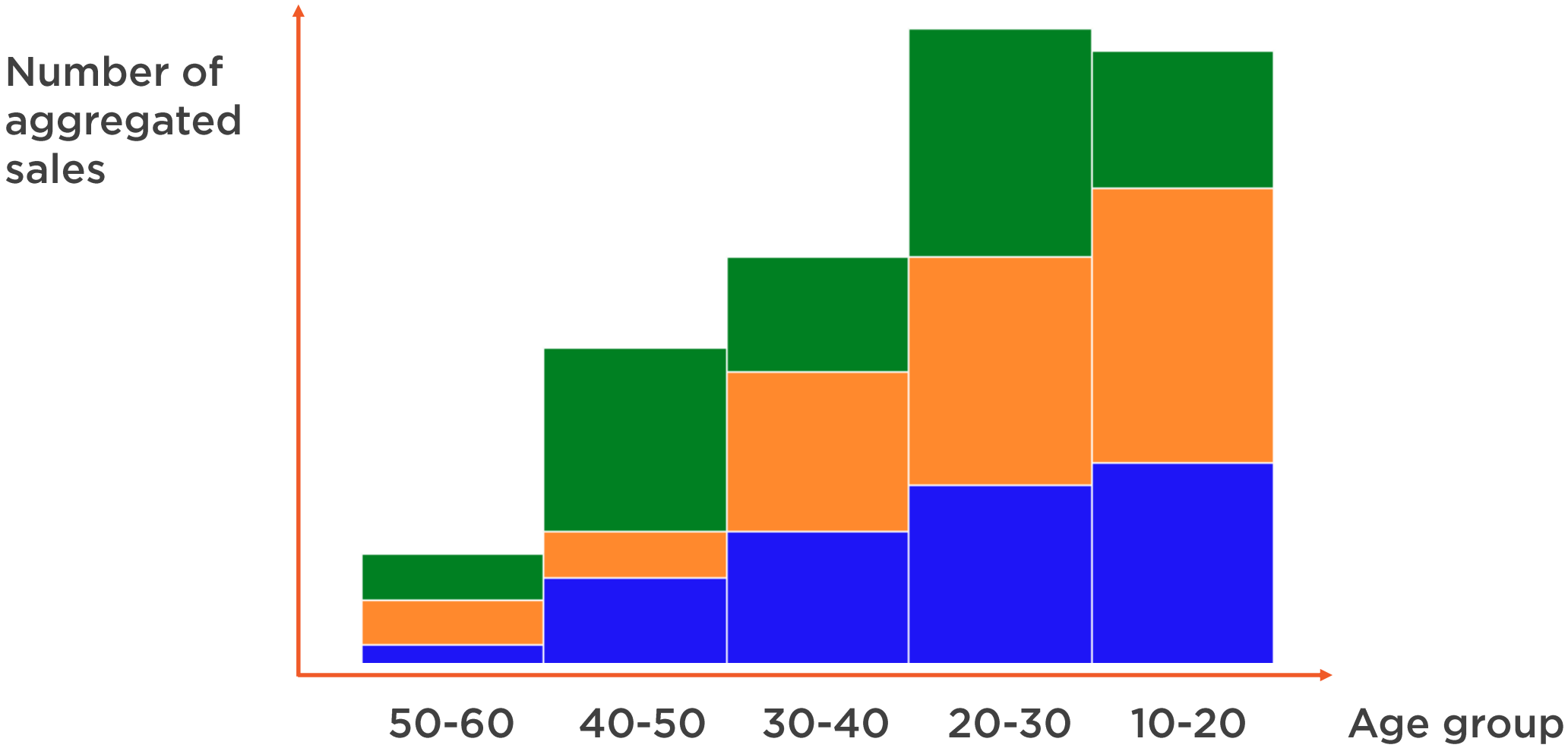
Stacked Area



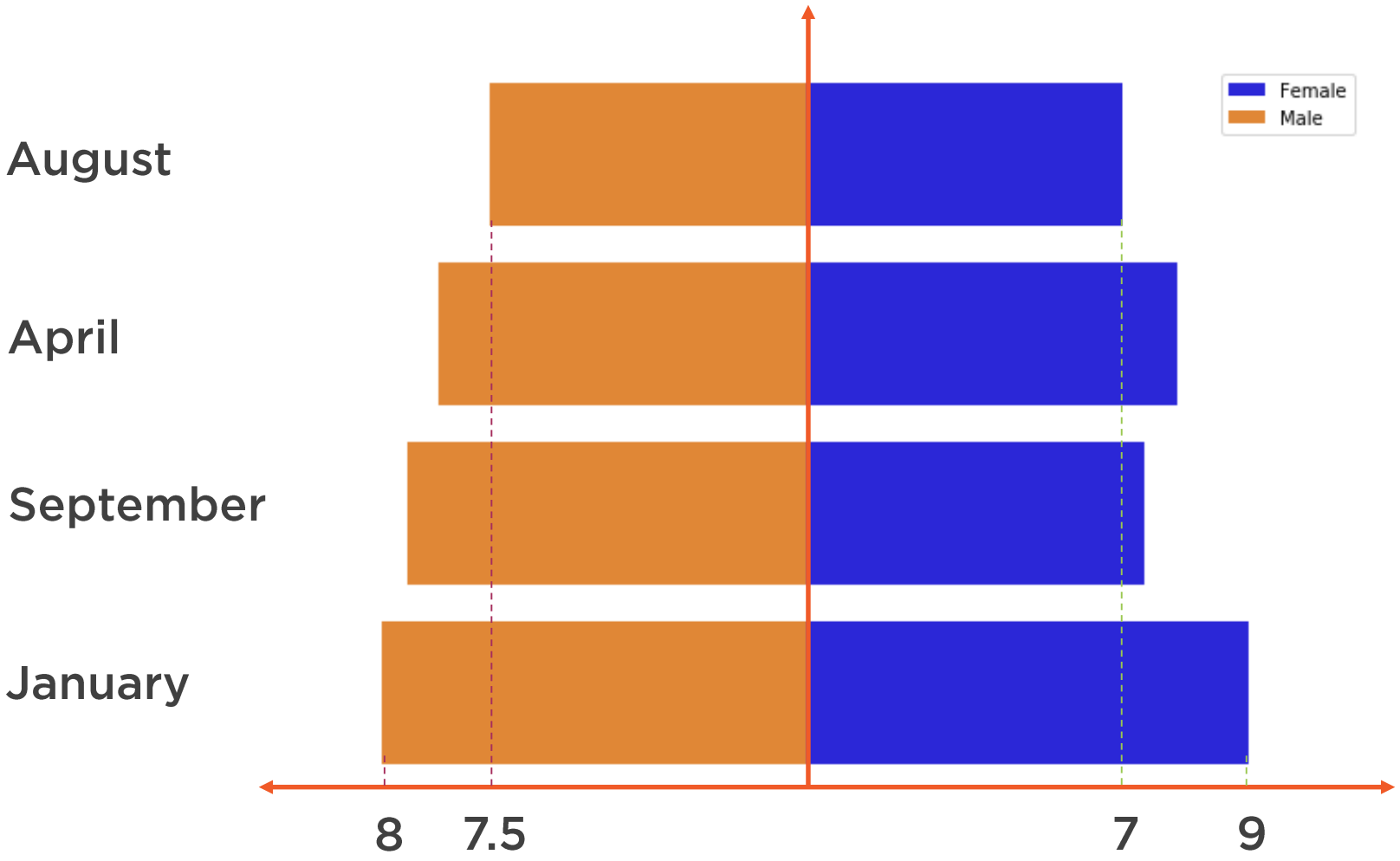
Stacked Area



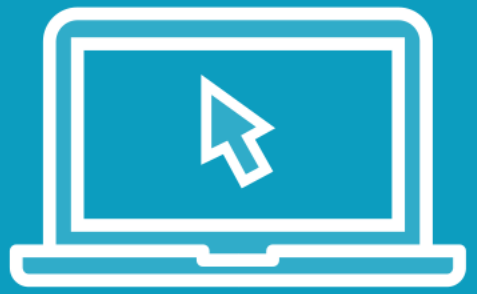
Stacked Bars



Tornado Chart



Demo



Learn how to plot multivariate composition charts with Python

Using Python packages

- Matplotlib
- Seaborn

Learn how to customize some simple graph aesthetics