```
if awake:
    code()
elif tired:
    drink_coffee()
```

Source

Welcome to Week 3 Lecture 1!

Data Science in Python & Machine Learning

Stack 1 Belt Exam

- Eligibility:
  - 90 % of Week 1 & 2 Assignments
    - **Including Resubmits**
    - Due: 9 AM PST on Thursday (02/10/22)

- Unlock Codes:
  - Thursday after Class
  - Must start exam by 11:59 pm Saturday

Note: You have 24 hours to Complete Exam

Week 3 CORE Assignments

These MUST be submitted by <u>Sunday February 13th</u>:

1) Average Height Exercise (Core)
2) Histograms & Boxplots (Core)
3) Project 1 - Part 3 (Core)

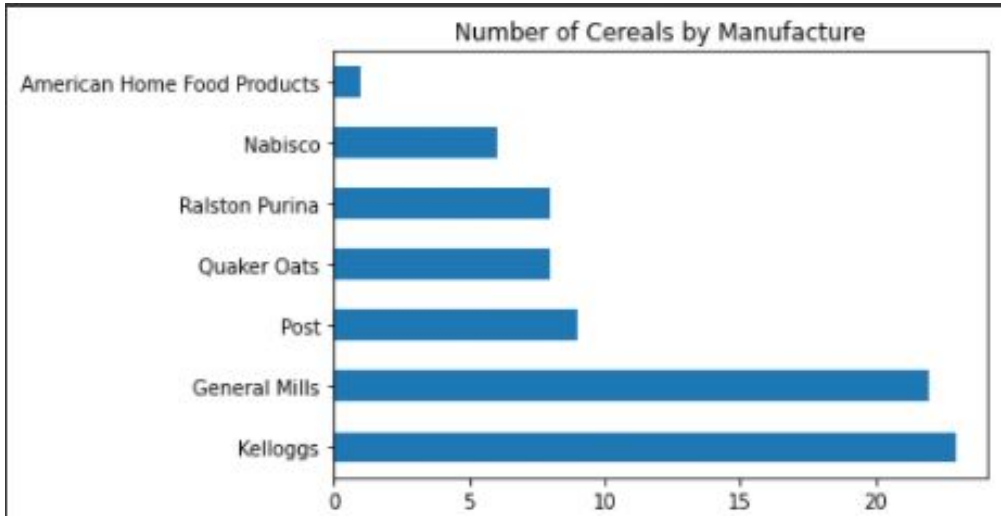
Learning Goals

By the end of this lesson you will be able to:

1. Univariate Data Visualizations
2. Exploratory Data Analysis
3. Use Matplotlib.pyplot
4. Create a barchart, histogram and boxplot

Univariate Data Visualizations

Exploring one column at a time



Exploratory Data Analysis

For You, the Data Scientist to Understand Your Data

(And Maybe for **Technical** Audiences)

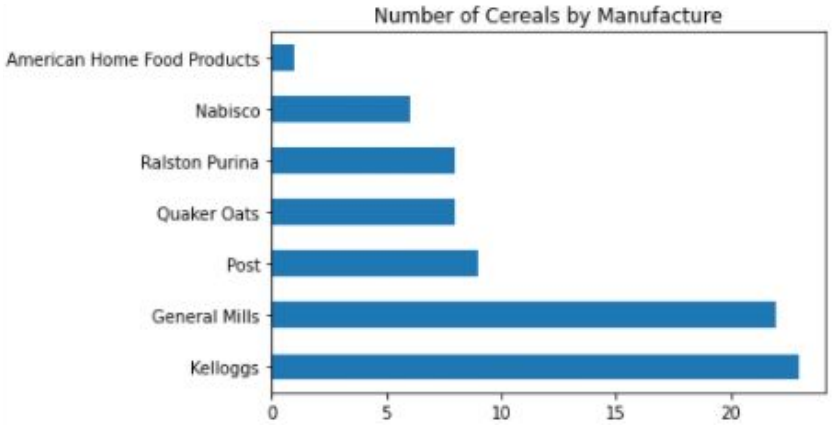
Exploratory Data Analysis

1. What is your data *like?*
    a. .shape
    b. .describe()

Exploratory Data Analysis

1. What is your dataset *like?*
   a. .shape
   b. .describe()

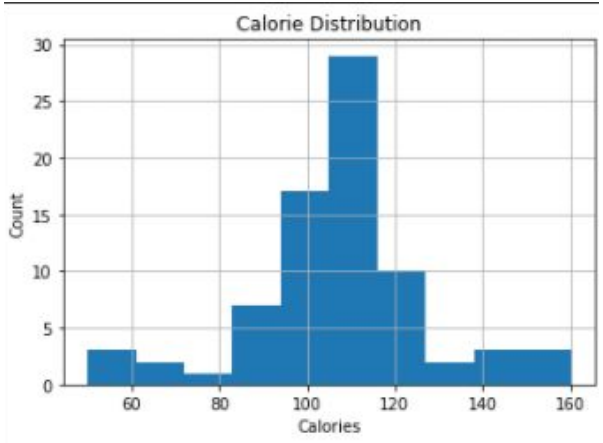
2. How is data *distributed?*
   a. Categorical
      i. How many of each category?



Exploratory Data Analysis

1. What is your dataset *like?*
   a. .shape
   b. .describe()

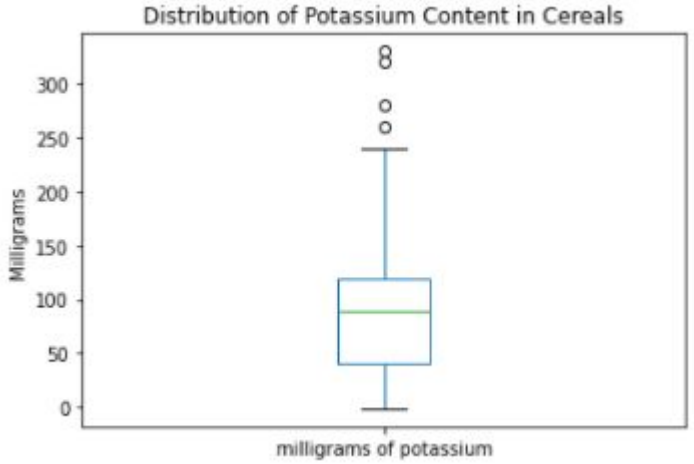
2. How is data *distributed?*
   a. Categorical
      i. How many of each category?
   b. Numeric
      i. What values are most of the data near?



Exploratory Data Analysis

1. What is your dataset *like?*
   a. .shape
   b. .describe()

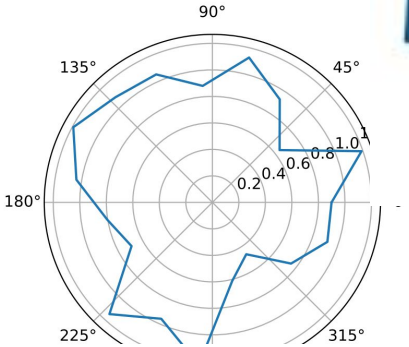
2. How is data *distributed?*
   a. Categorical
      i. How many of each category?
   b. Numeric
      i. What values are most of the data near?
      ii. Where are the outliers?

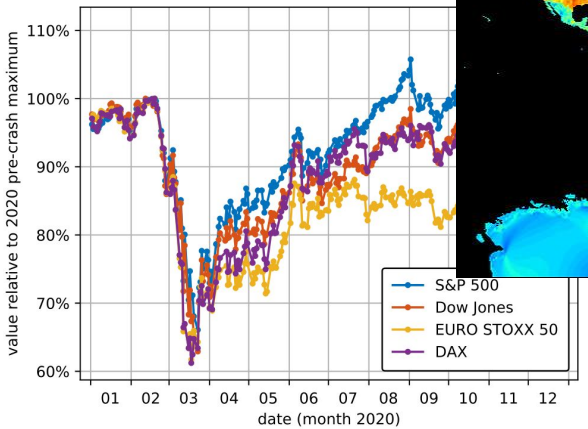


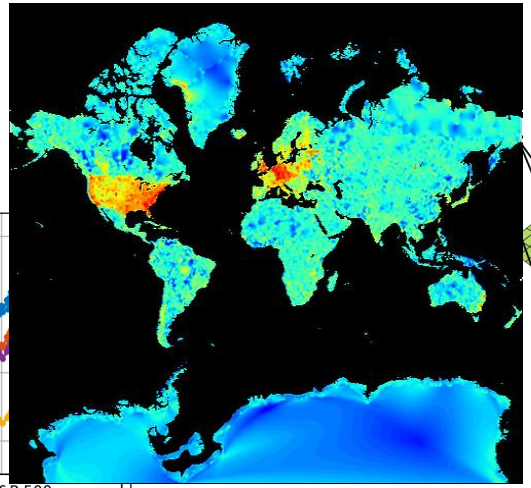
Univariate Analysis in Python

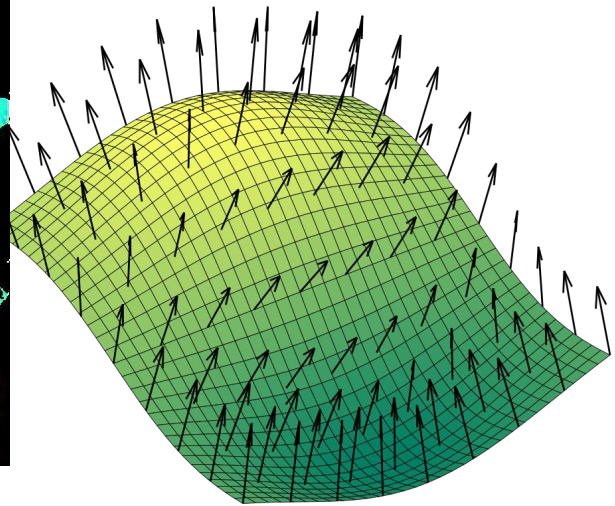
Exploring one variable at a time

matplotlib


90°
135°
45°
180°
225°
315°
0.2
0.4
0.6
0.8
1.0


value relative to 2020 pre-crash maximum
110%
100%
90%
80%
70%
60%
01
02
03
04
05
06
07
08
09
10
11
12
date (month 2020)
S&P 500
Dow Jones
EURO STOXX 50
DAX





The Power behind the Plotting

- Matplotlib.pyplot (plt)
- Seaborn (sns)
- Pandas Plotting

Sample Code!

In this sample univariate plotting notebook

Codealong:
Python Visualizations 101 - Part 1

- I've saved the traditional class activity to be an optional bonus assignment to practice your data cleaning & visualization skills.
  - See the final slide for the link.

- **Instead, we will walk through how to create & customize a specific visualization in 4 different ways**
  - Colab Notebook

And now for the FUN!

The one and only...

Clean and Plot Univariate Challenge!

Also Required:
Medical Data

Please share your best visualization with the class on Discord!