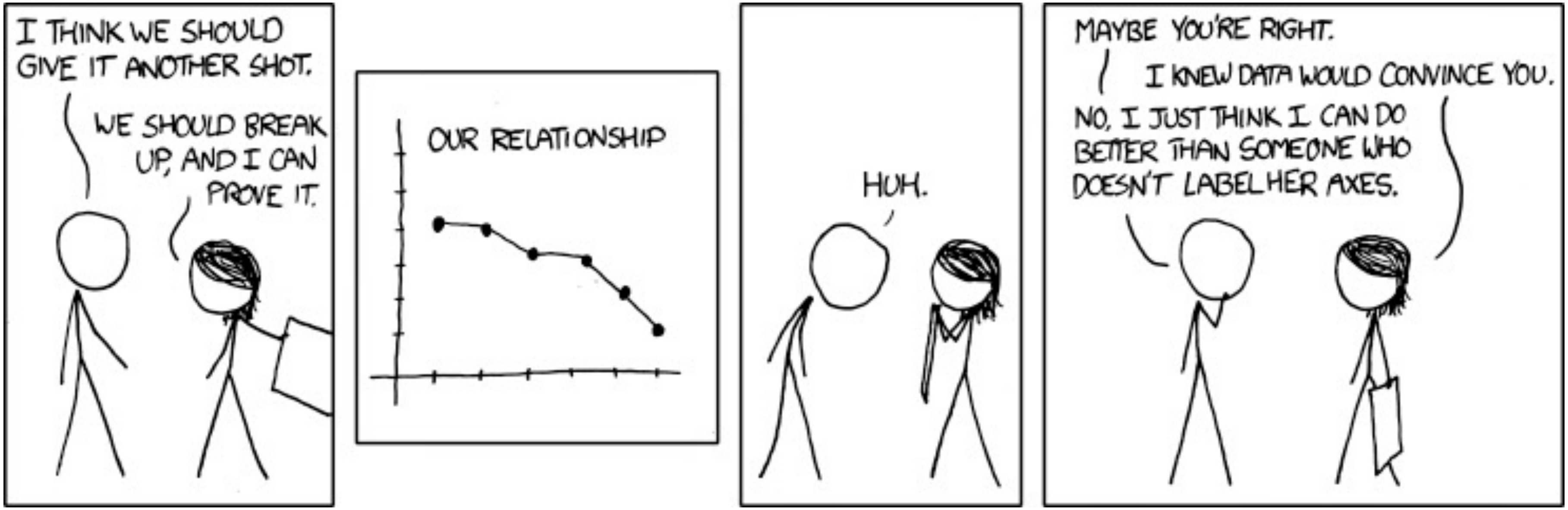
Today's question


I THINK WE SHOULD GIVE IT ANOTHER SHOT.
WE SHOULD BREAK UP, AND I CAN PROVE IT.
OUR RELATIONSHIP
HUH.
MAYBE YOU'RE RIGHT.
I KNEW DATA WOULD CONVINCE YOU.
NO, I JUST THINK I CAN DO BETTER THAN SOMEONE WHO DOESN'T LABEL HER AXES.

How's my data?

Data Analysis

Lab 8

Today's docket

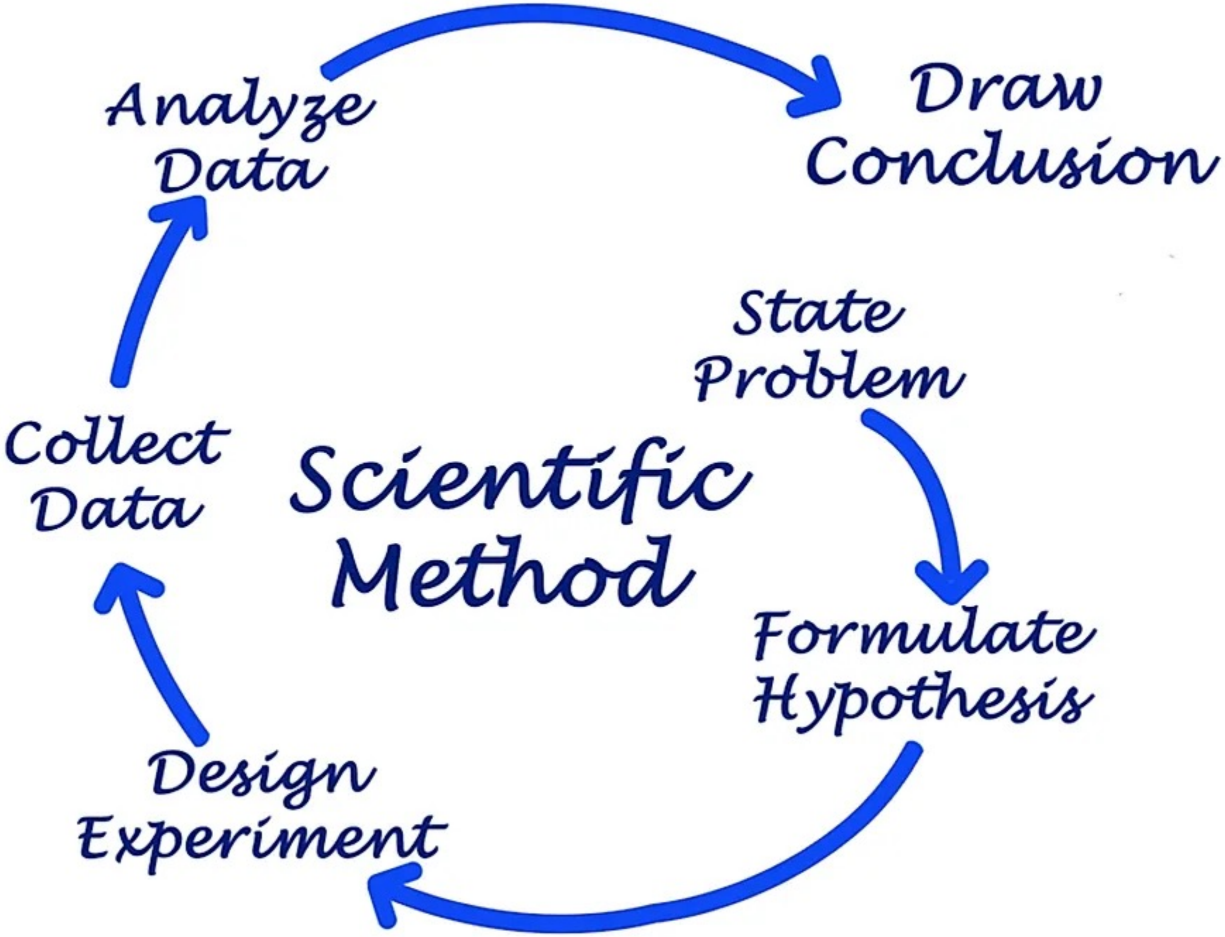
1. Data analysis

   Preprocessing, main analysis, visualization

2. Autism dataset

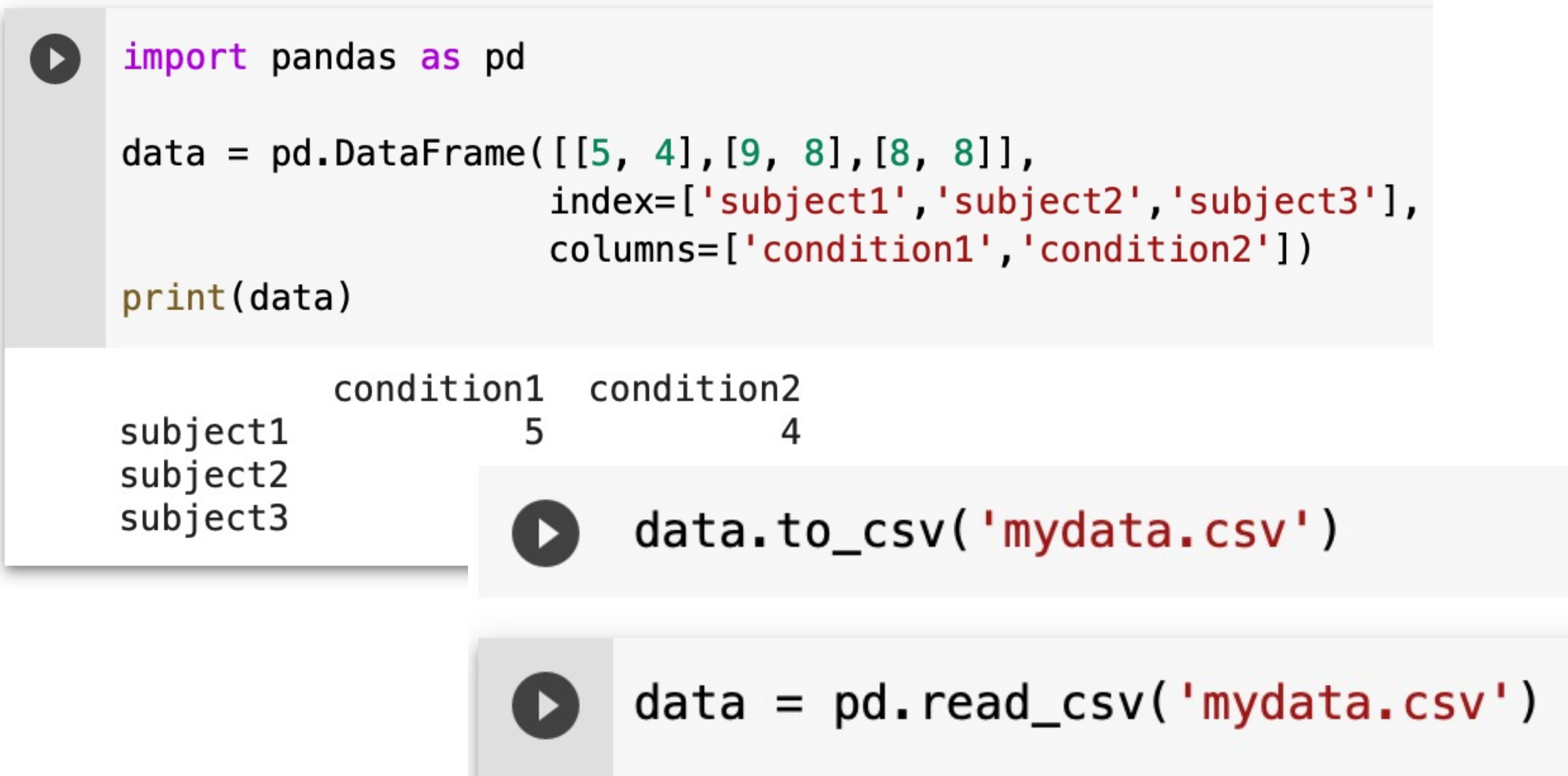
   Breakout session

Data analysis



Science as an ongoing process

Preprocessing



```
import pandas as pd

data = pd.DataFrame([[5, 4],[9, 8],[8, 8]],
                    index=['subject1','subject2','subject3'],
                    columns=['condition1','condition2'])
print(data)
```

```
          condition1  condition2
subject1           5           4
subject2
subject3
```

```
data.to_csv('mydata.csv')
```

```
data = pd.read_csv('mydata.csv')
```

Transforming the raw data into an understandable format

Preprocessing

- Data cleaning
  (removing incorrect and incomplete data, replacing missing values)
- Data integration
  (combining multiple sources into a single dataset)
- Data reduction
  (making the analysis easier, e.g., dimensionality reduction)
- Data transformation
  (changing the format or structure, e.g., smoothing, normalization)

Transforming the raw data into an understandable format

Main analysis

- Type: Inferential analysis, where conclusions drawn from the sample are inferred to apply to the larger population

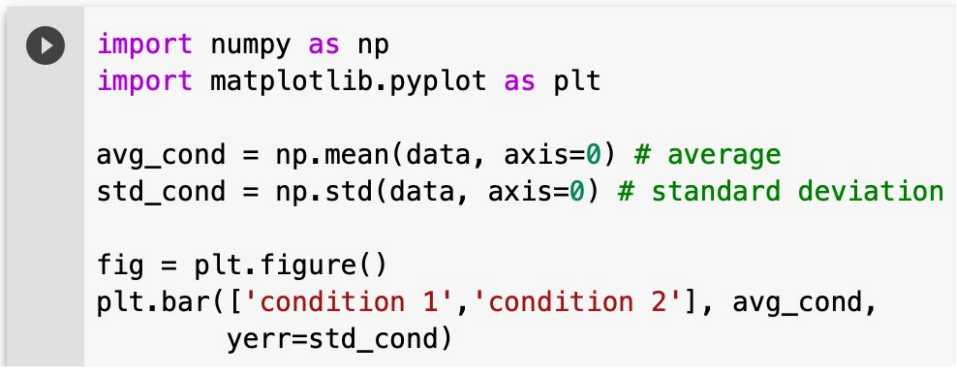
- Methods: *comparison* tests (e.g., *t*-test, ANOVA), *correlation* tests (e.g., Pearson), and *regression* tests (e.g., multiple linear regression)

- Focus: Reliability and validity (consistency and accuracy of observations)

Correspondence of observations to the conclusions

Visualization

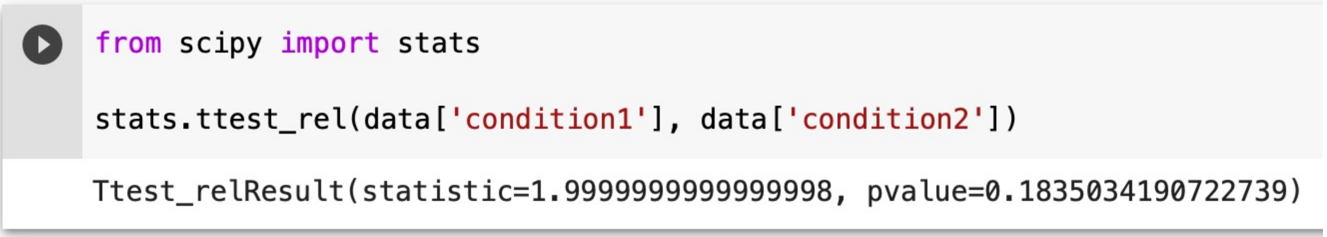
- Python libraries, including Matplotlib, Seaborn, Plotly, etc.



```
import numpy as np
import matplotlib.pyplot as plt

avg_cond = np.mean(data, axis=0) # average
std_cond = np.std(data, axis=0) # standard deviation

fig = plt.figure()
plt.bar(['condition 1','condition 2'], avg_cond,
        yerr=std_cond)
```



```
from scipy import stats

stats.ttest_rel(data['condition1'], data['condition2'])
```

```
Ttest_relResult(statistic=1.9999999999999998, pvalue=0.1835034190722739)
```

Identification and communication of patterns and trends in data

Today's docket

1. Data analysis

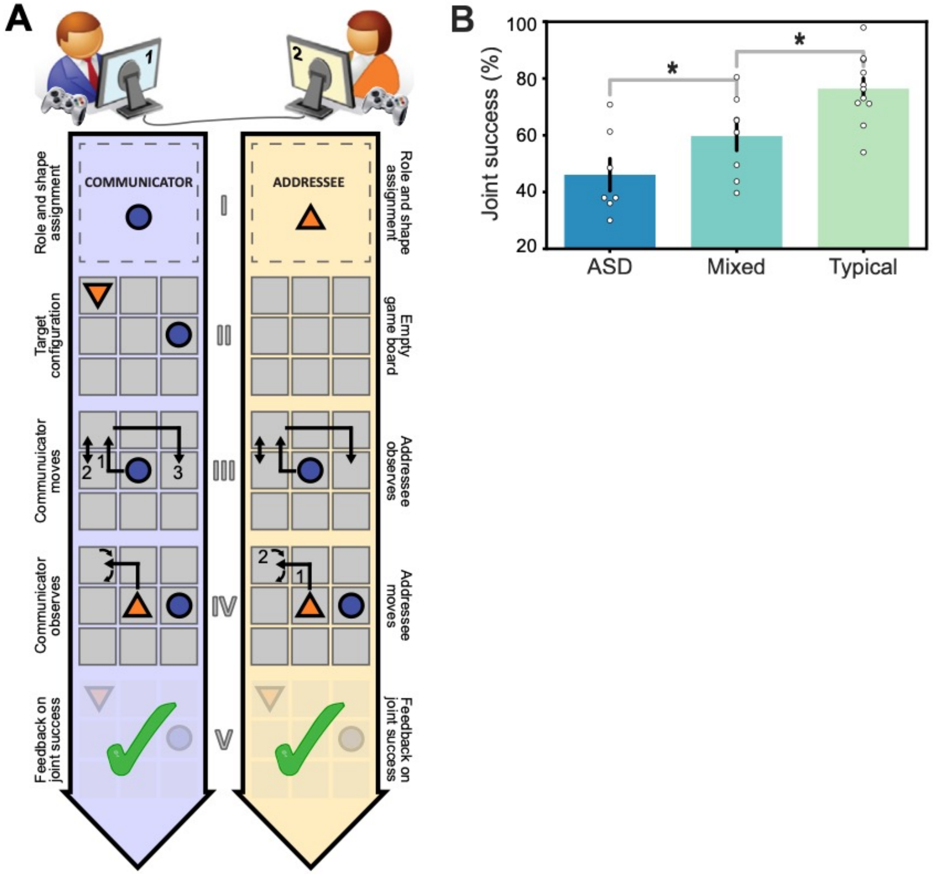
   Preprocessing, main analysis, visualization

2. Autism dataset

   Breakout session

Autism dataset

- Lab8_TCG_ASD.ipynb



Take-home concepts

- Data analysis is about applying statistical and/or logical techniques to describe, illustrate, and evaluate observations

Next up

- *Wrap up Data Collection asap*
- *Start Data Analysis*
- *Hackathon on Wednesday (R-hour)*
- *Presentations on Friday*