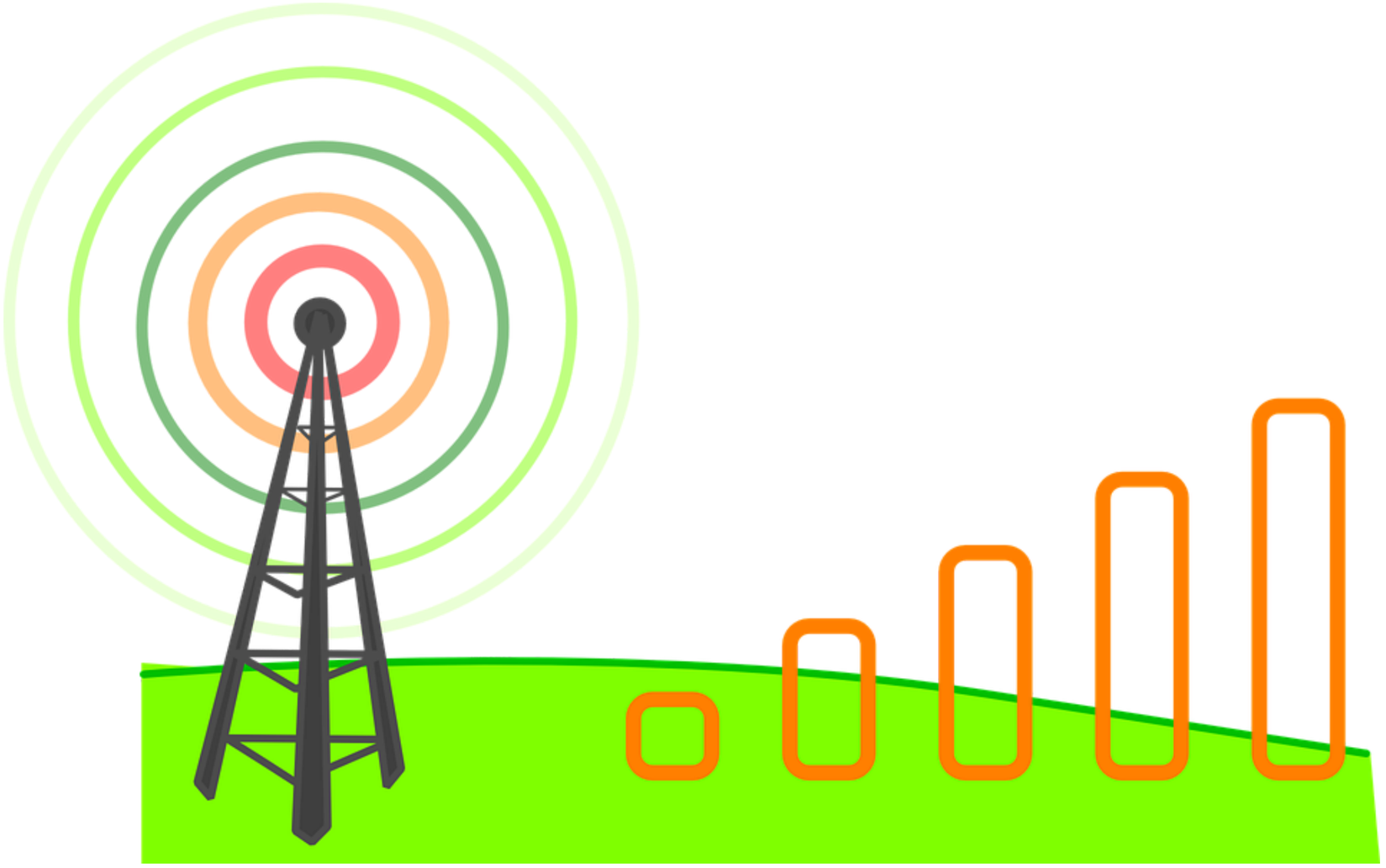
Making a scatter plot

INTRODUCTION TO DATA SCIENCE IN PYTHON

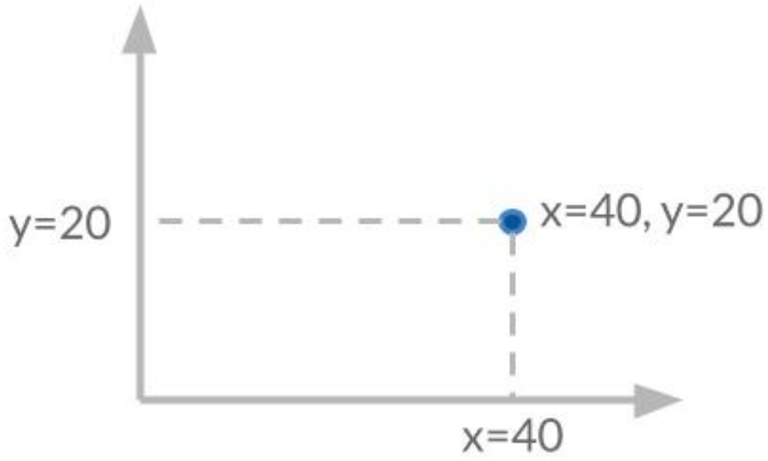
Hillary Green-Lerman
Lead Data Scientist, Looker

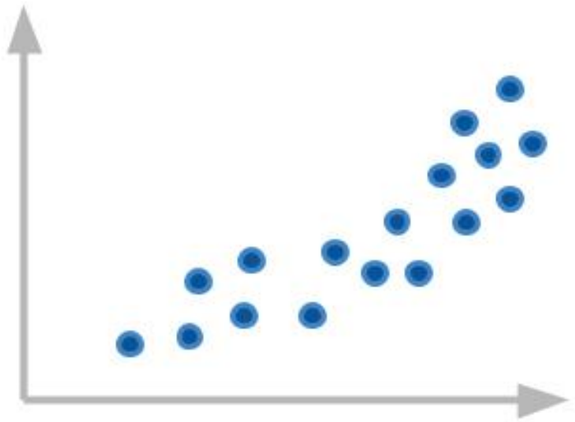
datacamp

Mapping Cell Phone Signals

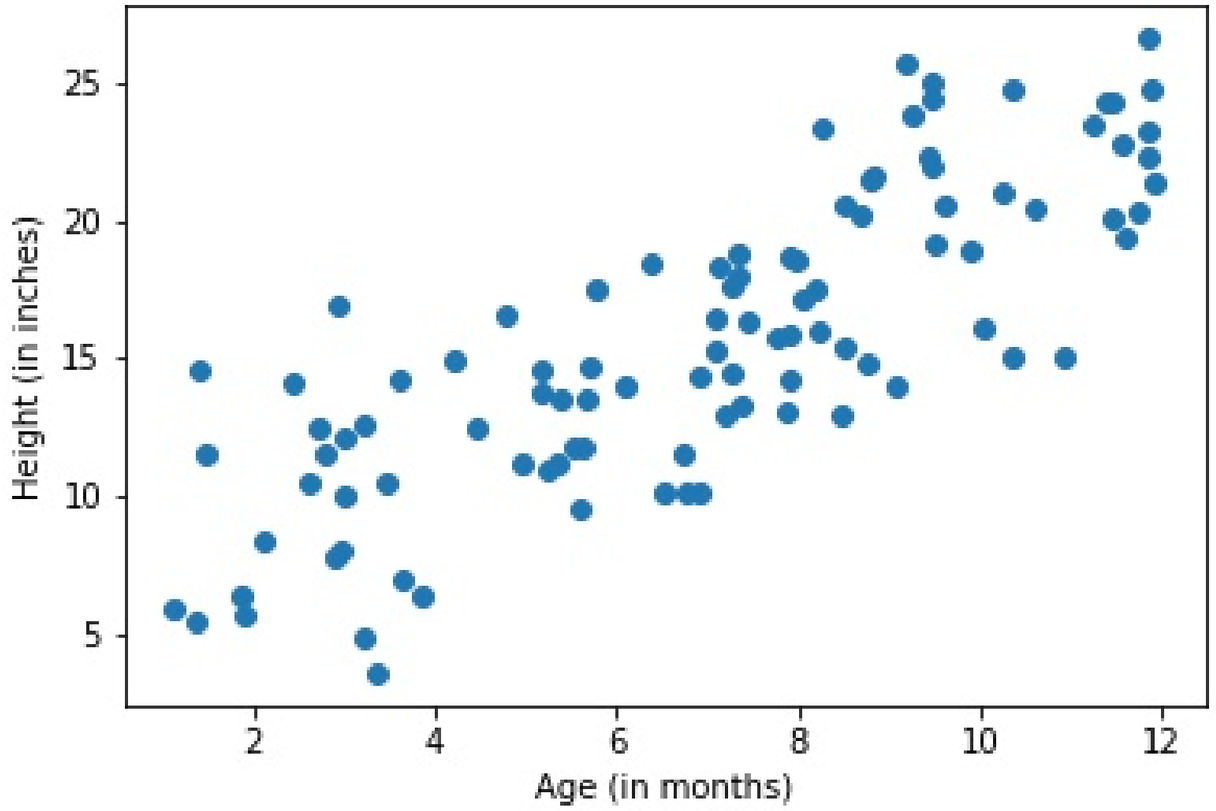


What is a scatter plot?





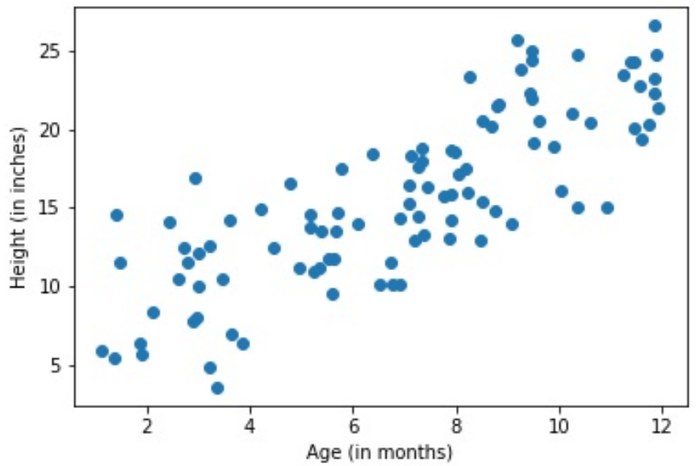
What is a scatter plot?



Creating a scatter plot

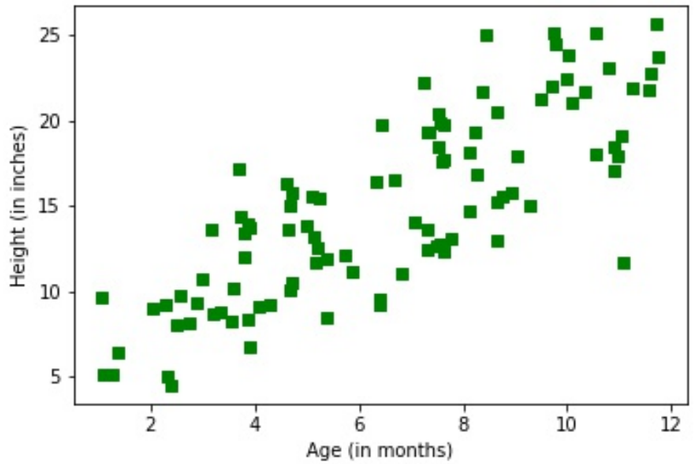
```
plt.scatter(df.age, df.height)
```

```
plt.xlabel('Age (in months)')
plt.ylabel('Height (in inches)')

plt.show()
```

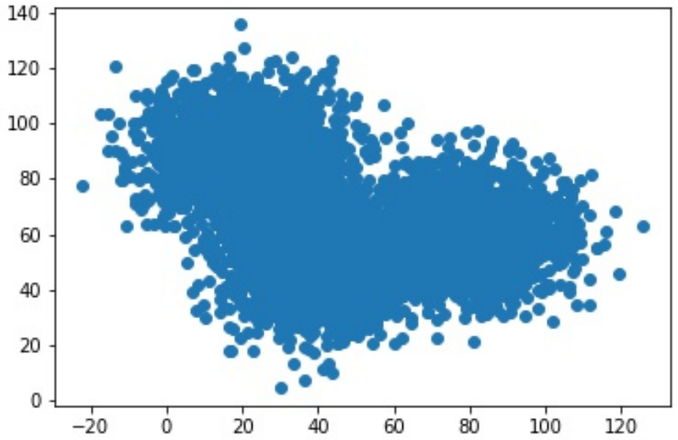


Keyword arguments

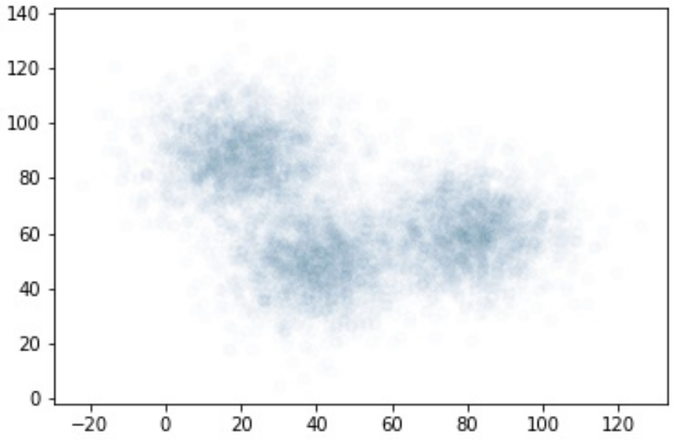
```
plt.scatter(df.age, df.height,
            color='green',
            marker='s')
```



Changing marker transparency




```
plt.scatter(df.x_data,
            df.y_data,
            alpha=0.1)
```



Let's practice!

INTRODUCTION TO DATA SCIENCE IN PYTHON

Making a bar chart

INTRODUCTION TO DATA SCIENCE IN PYTHON

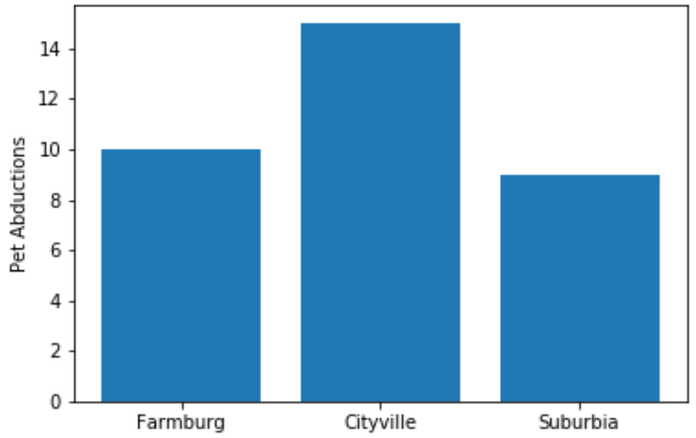
Hillary Green-Lerman
Lead Data Scientist, Looker

datacamp

Comparing pet crimes

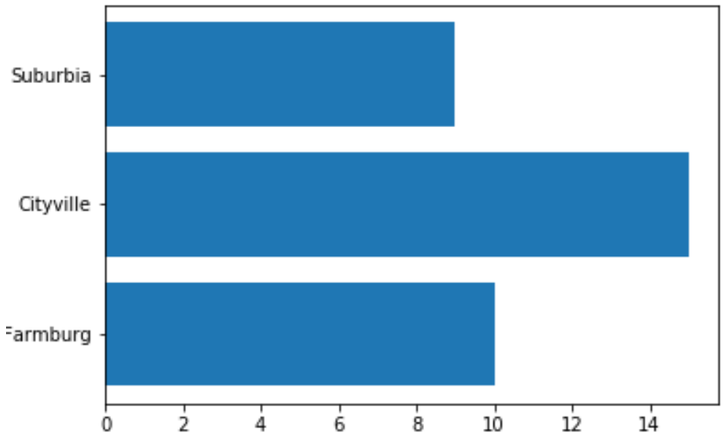
precinct	pets_abducted
Farmburg	10
Cityville	15
Suburbia	9

```
plt.bar(df.precinct,
        df.pets_abducted)

plt.ylabel('Pet Abductions')
plt.show()
```



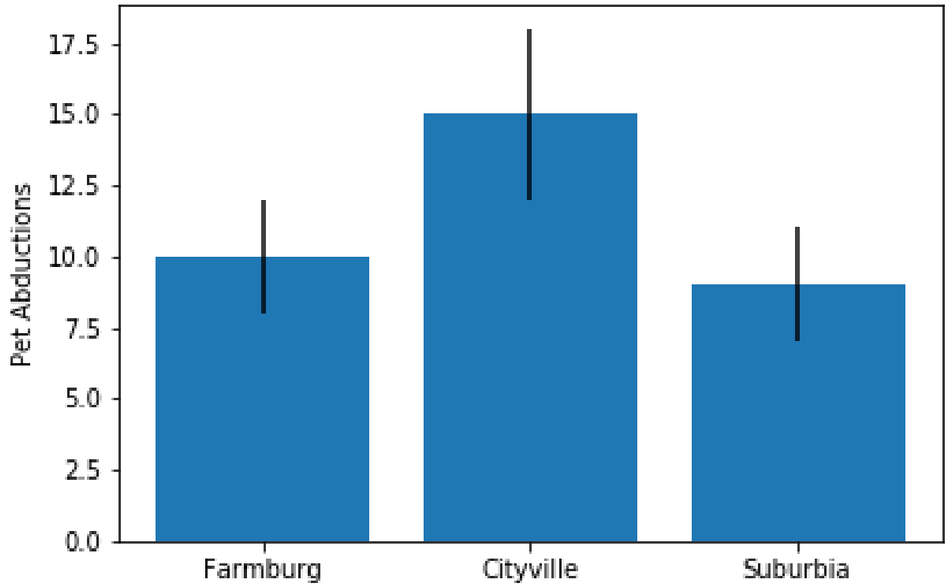
Horizontal bar charts

```
plt.barh(df.precinct,
         df.pets_abducted)

plt.ylabel('Pet Abductions')
plt.show()
```

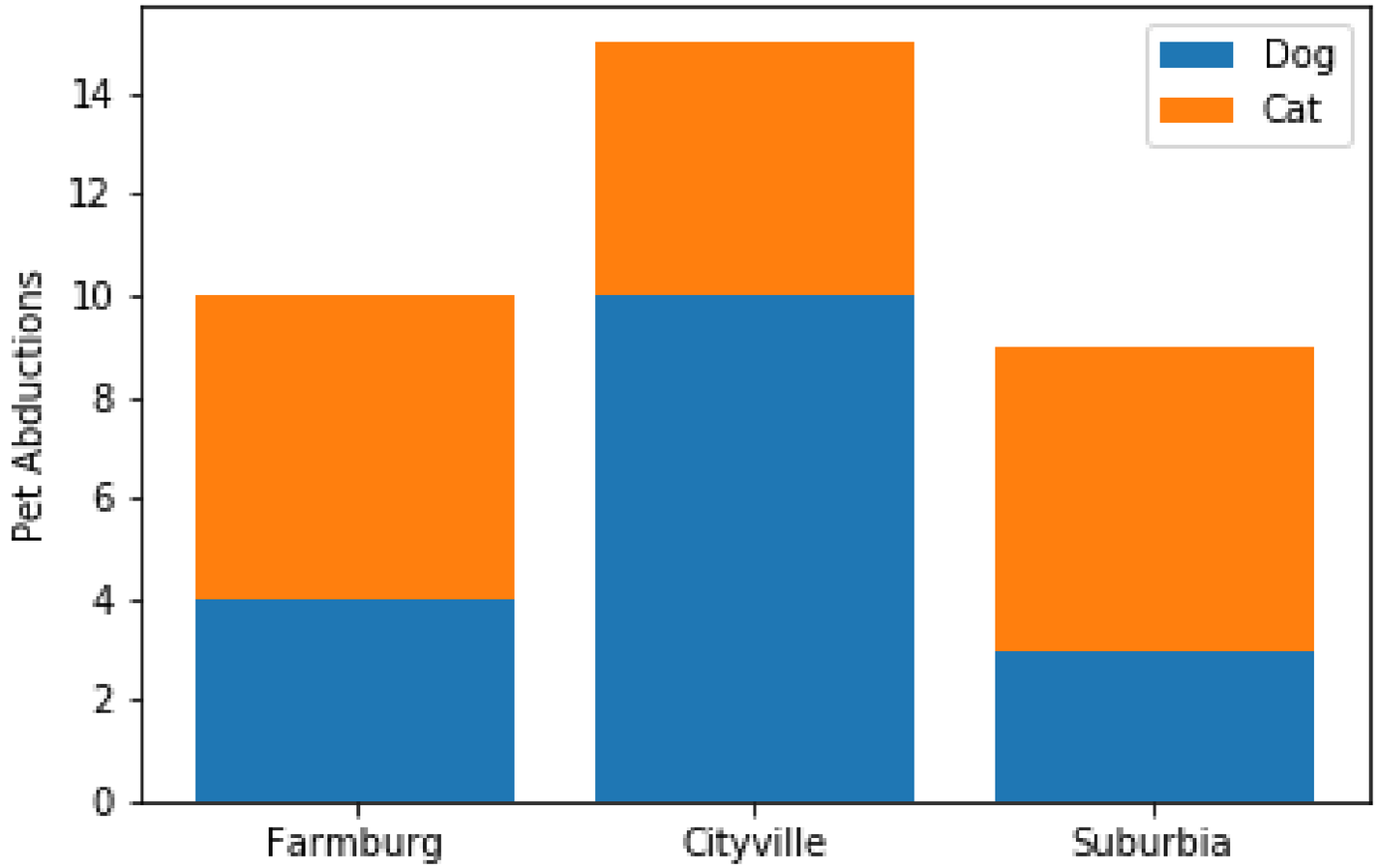


Adding error bars

```
plt.bar(df.precinct, df.pet_abductions,
        yerr=df.error)

plt.ylabel('Pet Abductions')
plt.show()
```



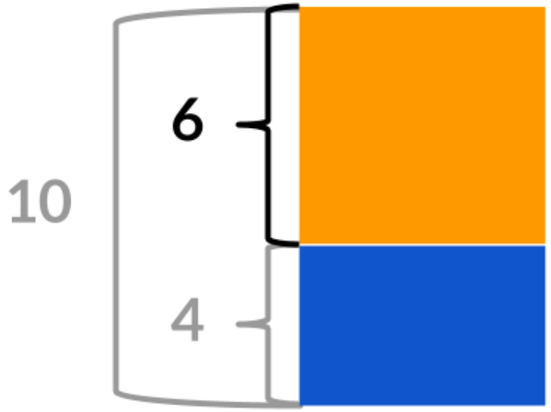
Stacked bar charts



Stacked bar charts


```
plt.bar(df.precinct, df.dog)
```

Stacked bar charts

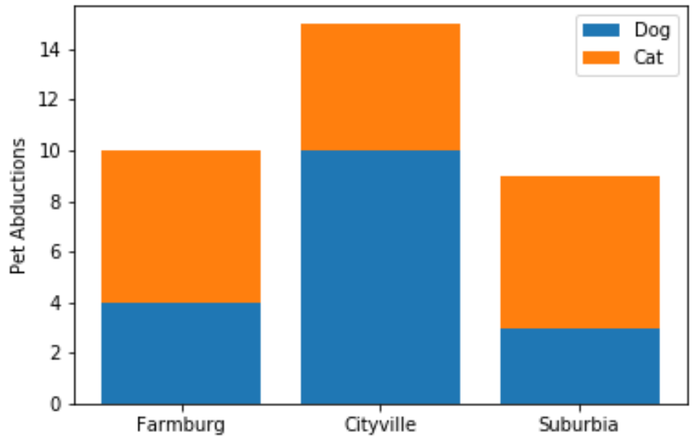



```
plt.bar(df.precinct, df.cat,
bottom=df.dog)
```

```
plt.bar(df.precinct, df.dog)
```

Stacked bar charts

```
plt.bar(df.precinct, df.dog,
        label='Dog')

plt.bar(df.precinct, df.cat,
        bottom=df.dog,
        label='Cat')

plt.legend()
plt.show()
```



Let's practice!

INTRODUCTION TO DATA SCIENCE IN PYTHON

Making a histogram

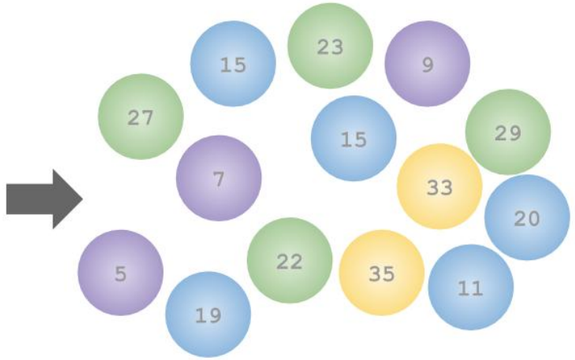
INTRODUCTION TO DATA SCIENCE IN PYTHON

Hillary Green-Lerman
Lead Data Scientist, Looker

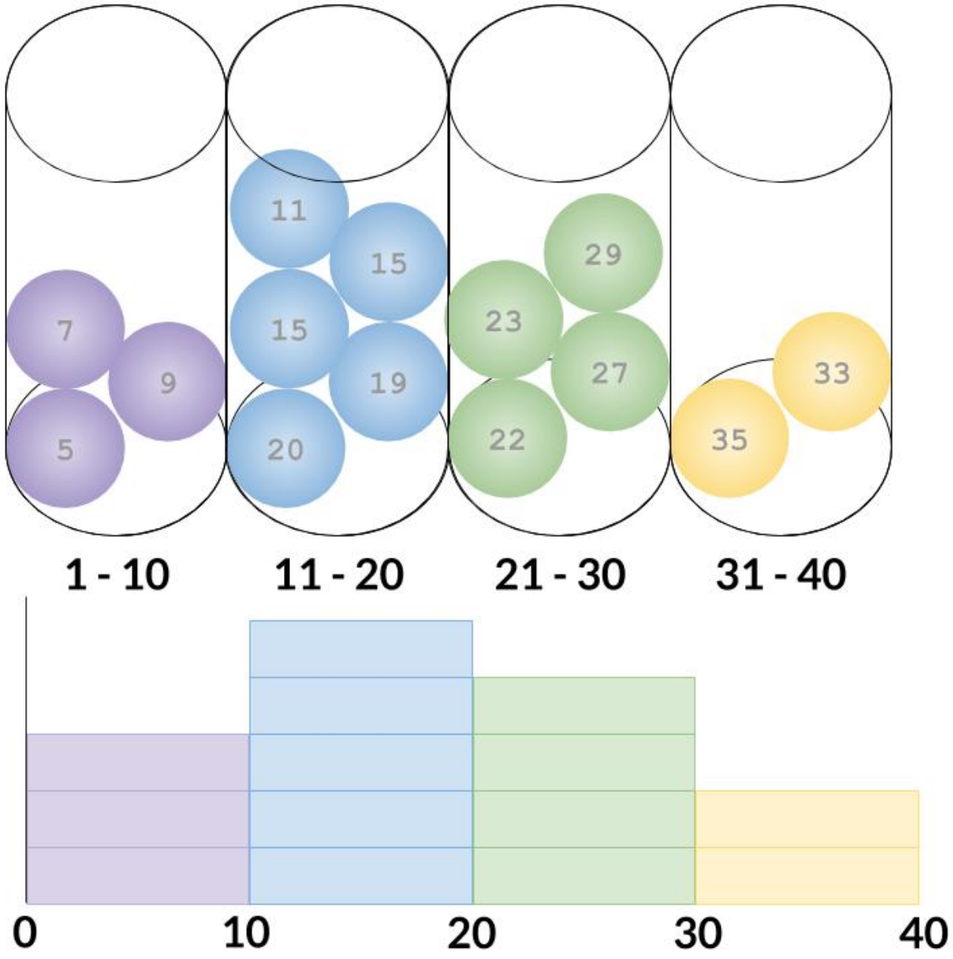
datacamp

Tracking down the kidnapper

Gravel Radius (mm)
5
17
7
20
42
35
21
...

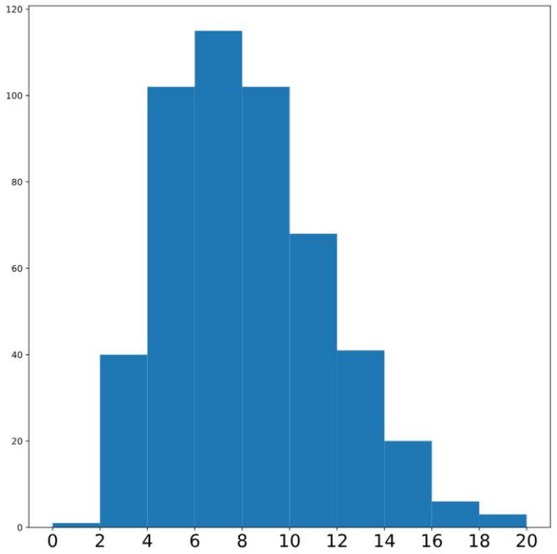


What is a histogram?



Histograms with matplotlib

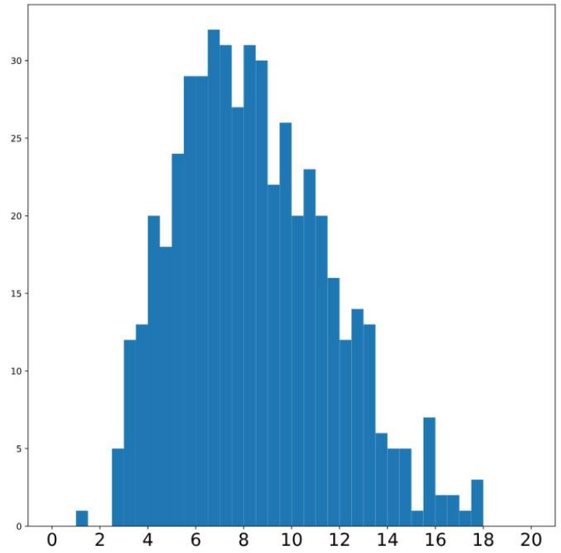
```
plt.hist(gravel.mass)


plt.show()
```



Changing bins

```
plt.hist(data, bins=nbins)
```

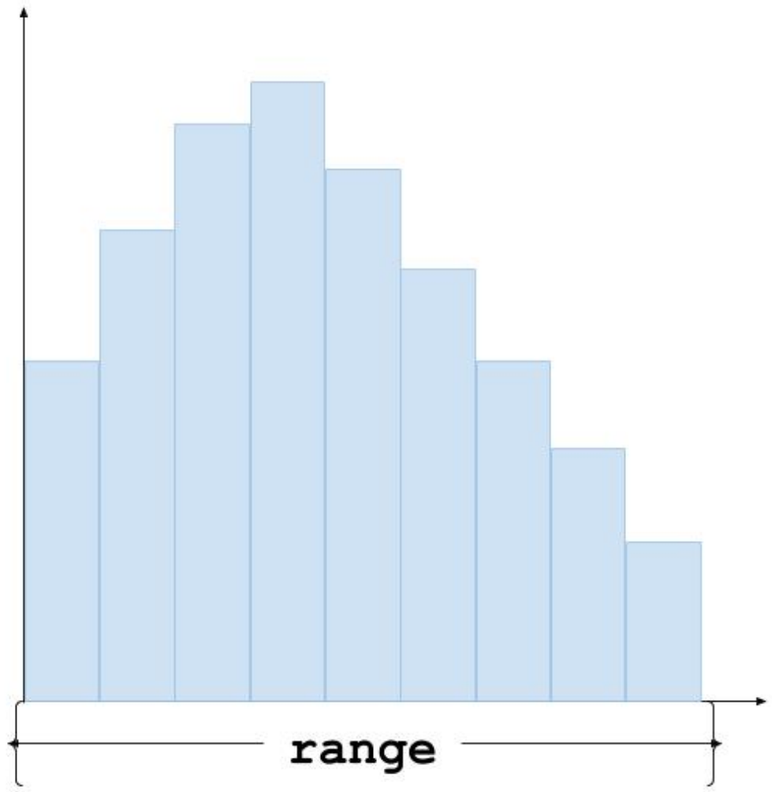
```
plt.hist(gravel.mass, bins=40)
```



Changing range

```
plt.hist(data,
         range=(xmin, xmax))
```

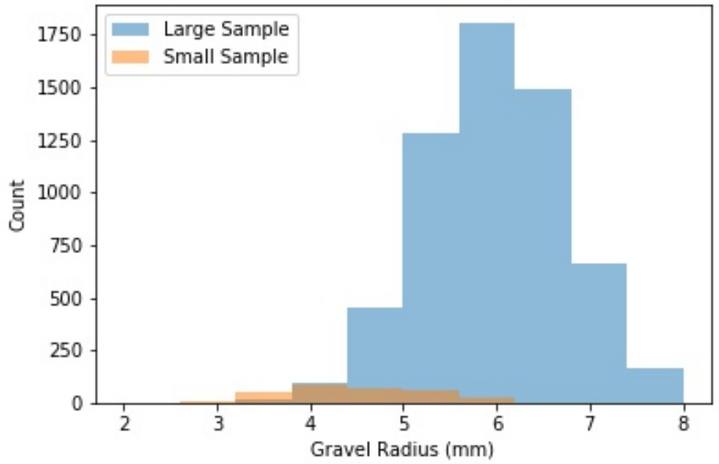
```
plt.hist(gravel.mass,
         range=(50, 100))
```



Normalizing

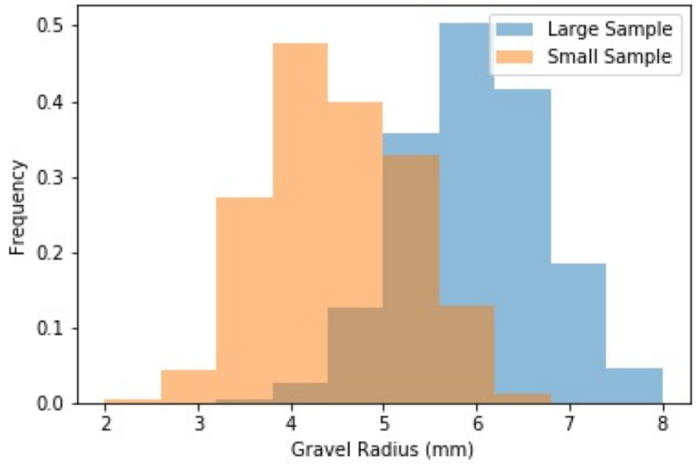
Unnormalized bar plot

```
plt.hist(male_weight)
plt.hist(female_weight)
```



Sum of bar area = 1

```
plt.hist(male_weight, density=True)
plt.hist(female_weight, density=True)
```



Let's practice!

INTRODUCTION TO DATA SCIENCE IN PYTHON

Recap of the rescue

INTRODUCTION TO DATA SCIENCE IN PYTHON

Hillary Green-Lerman
Lead Data Scientist, Looker

datacamp

You did it!

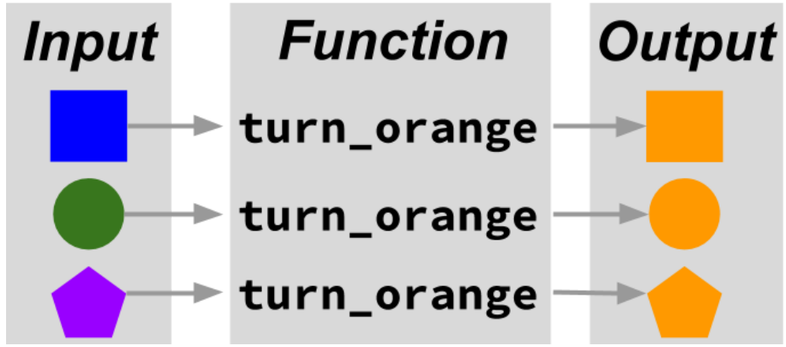
Modules and variables

- Modules group functions together
- Add a module using `import`
- `import` happens at the beginning of a script file
- Variables store data: strings or floats

```
import pandas as pd
import numpy as np
```

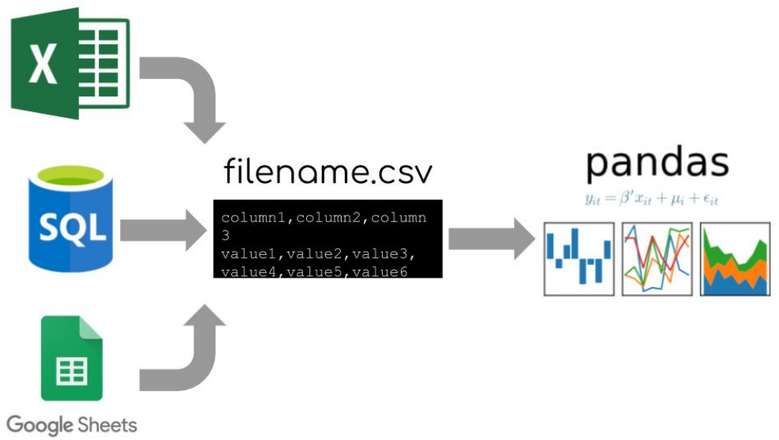
Using functions

- Perform a task
- Positional arguments
- Keyword arguments



Working with tabular data

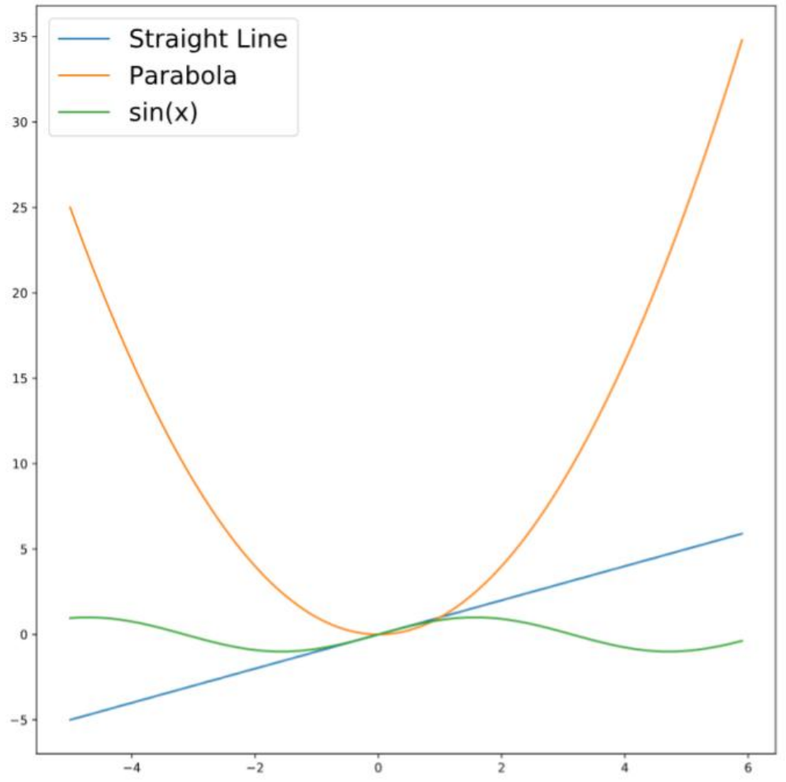
- `import pandas as pd`
- DataFrames store tabular data
- Inspect data using `.head()` or `.info()`
- Select rows using logic




```
credit_reports[
    credit_report.suspect ==
    'Freddy Frequentist']
```

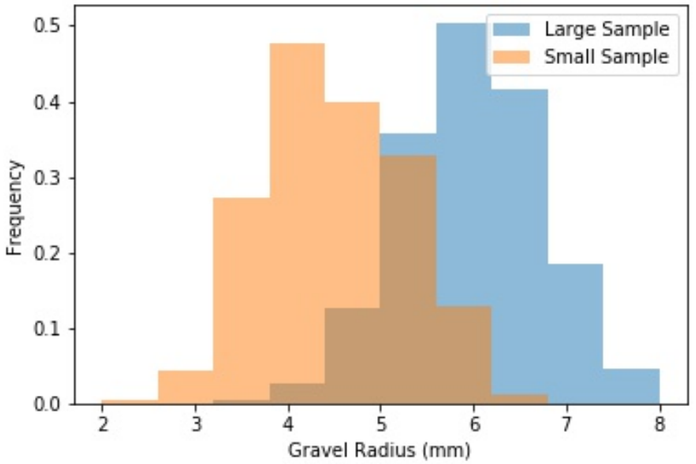
Creating line plots

- `from matplotlib import pyplot as plt`
- Use `plt.plot()` to create a line plot
- Modify line plots with keyword arguments
- Add labels and legends



More plot types

- `plt.scatter()` shows individual data points
- `plt.bar()` creates bar charts
- `plt.hist()` visualizes distributions



Great job!

INTRODUCTION TO DATA SCIENCE IN PYTHON