Exploring Data

Reindert-Jan Ekker

@rjekker   http://nl.linkedin.com/in/rjekker

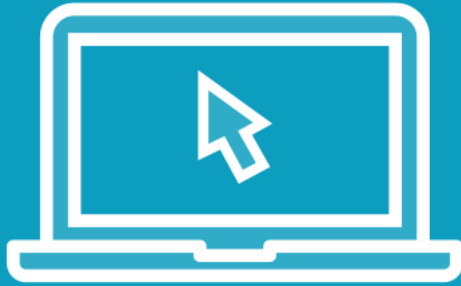# Overview

Exploring a new dataset

Numbers and types of observations

Overview of statistics

Simple plots

Demo

Properties of a Data Set

- Row and column counts

- Data types

- Previewing raw data

Demo

Statistic Exploration

- Statistical overview
- Other statistical methods
- Basic plotting

First Exploration Steps

```python
# How many rows and columns are there?

>>> df.shape

# Data type per column, memory usage

>>> df.info()

# Inspecting start and end of the data: head() and tail()

>>> df.head()

# An overview of basic statistics

>>> df.describe()
```

Statistical Properties

```python
# These work both on Series and DataFrames

# Maximum, minimum, mean, median

df.max(), df.min(), df.mean(), df.median()

# counting values

df.mode(), Series.value_counts()

# Distribution

df.std(), df.var()

df.skew(), df.kurtosis(), df.quantile()
```