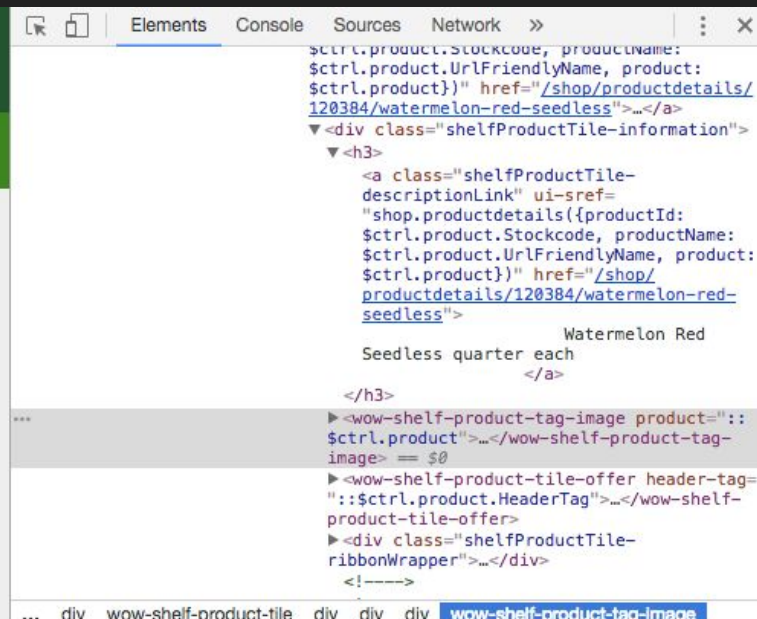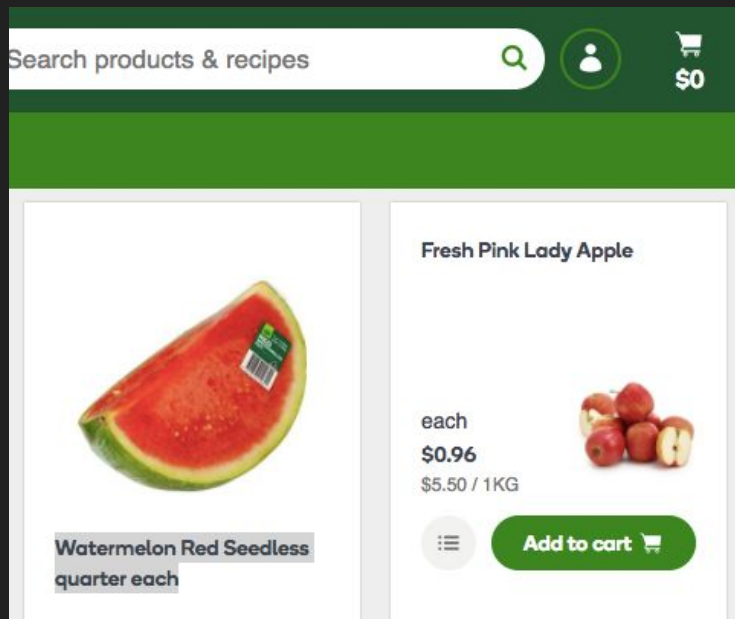# Comparing Grocery Prices

- Comparing prices between Woolworths and Coles
- Scrape data from web using Python and Javascript
- Use statistical tools to analyse data

Data sets

- Products are rendered dynamically using Javascript
- Selenium:
  - Library that lets you scrape JS rendered content
  - Inject Javascript, extract data, return to Python
  - Also used for bots, automation
- Save data as JSON files

```
{
    "category": "meat-seafood-deli",
    "href": "https://www.woolworths.com.a
    "imgName": "148292.jpg",
    "imgSrc": "https://cdn1.woolworths.me
    "name": "Macro Meats Kangaroo Mince 1
    "price": 9.99,
    "unitPrice": "$9.99 / 1KG"
},
```

```
{
    "brand": "Coles",
    "category": "meat-seafood-deli",
    "name": "3 Star Beef Mince",
    "price": "$8.00 per 1Kg",
    "special": "False",
    "subcategory": "fresh-meat"
},
```

Finding similar products

- ● Use scikit-learn to find products with similar names

```
==========================================
Similarity: 0.5347955590782588
Coles product: Greek Feta
Woolworths product: Dodoni Greek Barrel Aged Feta 200g
Coles price: $29.0 per 1.0 kg
Woolworths price: $34.75 per 1.0 kg
Difference: 5.75

==========================================
Similarity: 0.6135922078793593
Coles product: Crumbed Whiting Fillets
Woolworths product: Just Caught Whiting Fillets Crumbed 1kg
Coles price: $35.09 per 1.0 kg
Woolworths price: $10.0 per 1.0 kg
Difference: -25.0900000000000003

==========================================
```
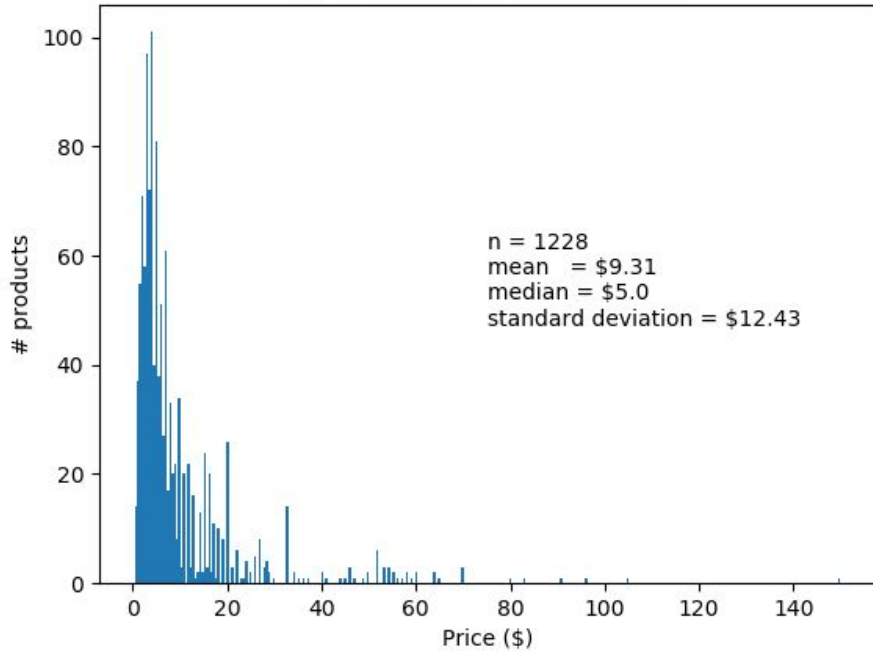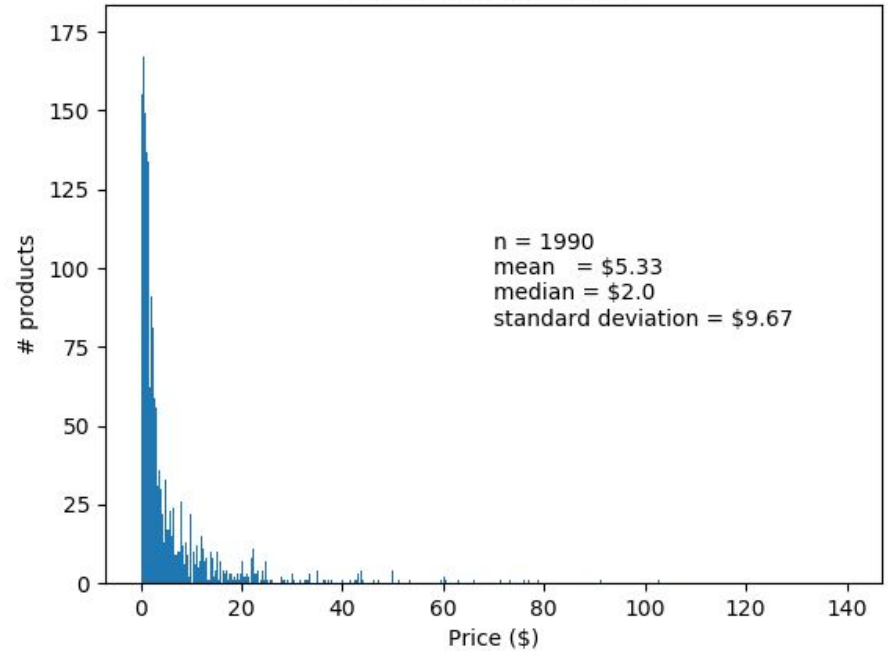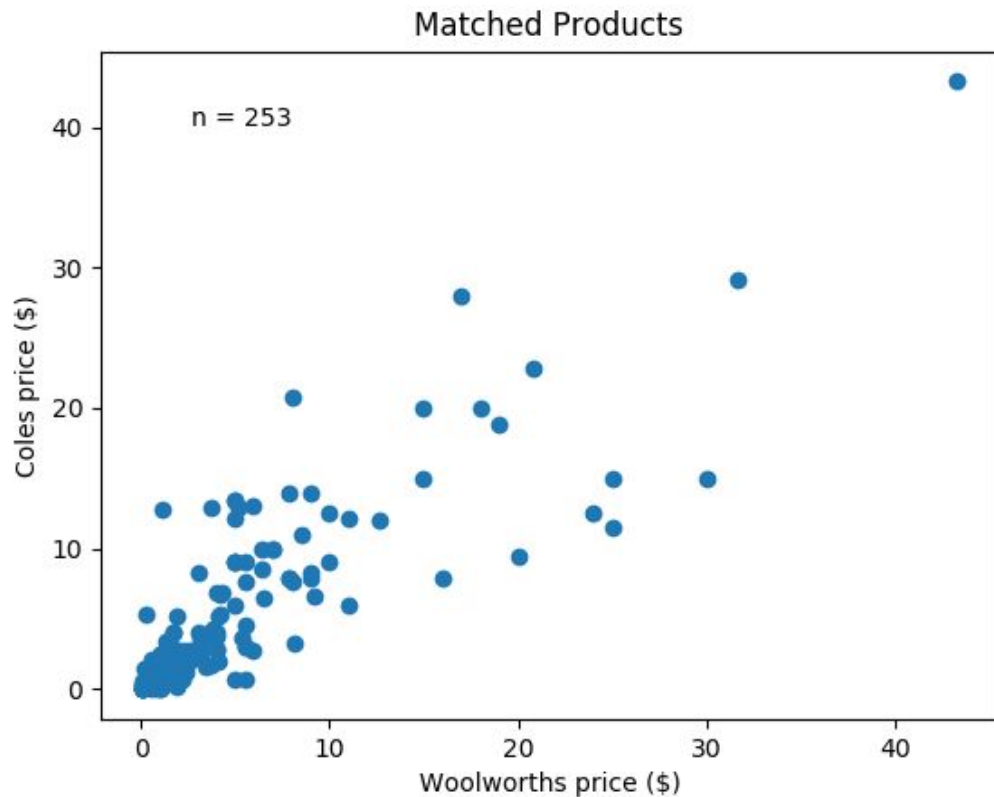
Visualisations

All Woolworths Product Prices

n = 1228
mean   = $9.31
median = $5.0
standard deviation = $12.43

All Coles Product Prices

n = 1990
mean   = $5.33
median = $2.0
standard deviation = $9.67
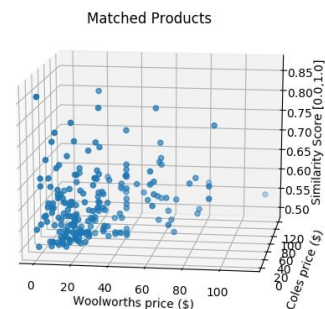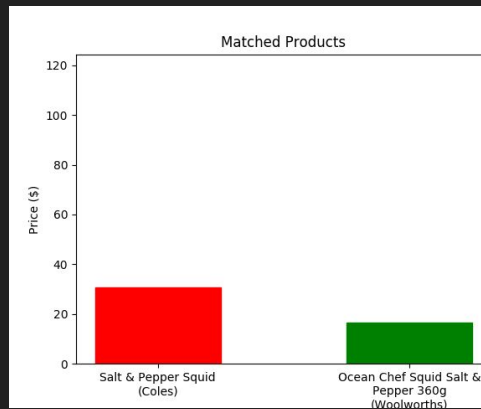
Visualisations
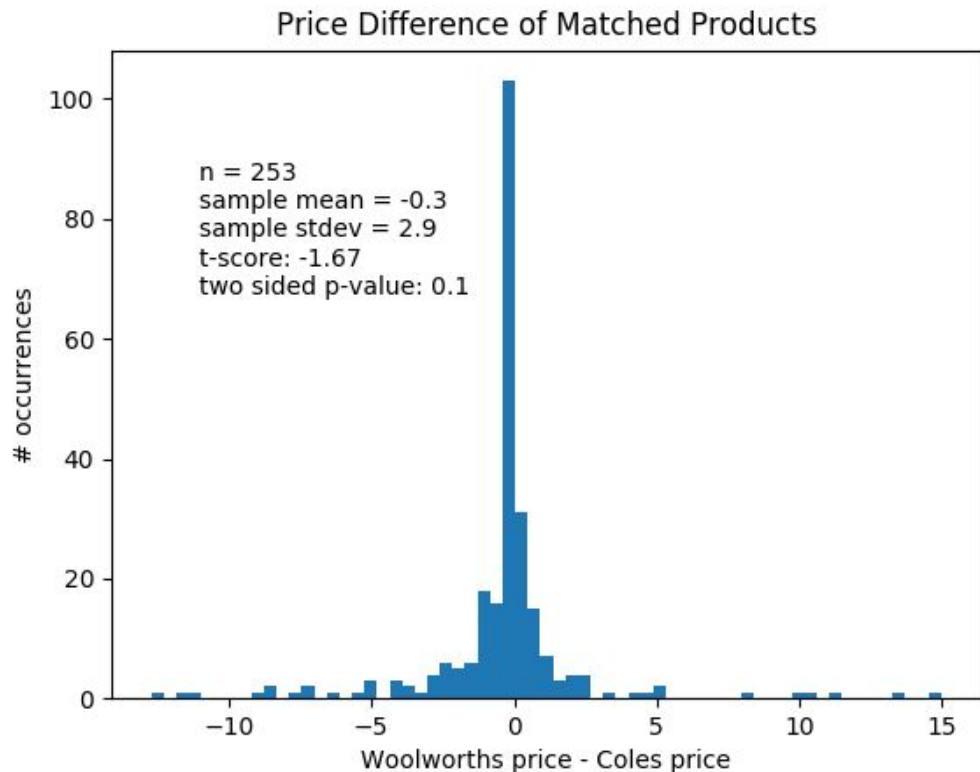

Animated bar plot and 3D plot

Analysis

- Scipy provides statistical tests
- Hypothesis test
- H0: μ difference = 0
- HA: μ difference != 0
- p-value = 0.96 > 0.05
- Fail to reject null hypothesis
- Price difference is not statistically significant
- We are not statisticians

Price Difference of Matched Products

n = 253
sample mean = -0.3
sample stdev = 2.9
t-score: -1.67
two sided p-value: 0.1

```
Number of similar products found: 253
Total difference in price, woolworths - coles: -76.98999999999995
Woolworths wins
t-score: -1.66768766254
two sided p-value: 0.0966198812861
p-value > 0.05
```