Assignment Scientific Methods

Filip Lobpreis

2024-10-23

We are loading the historical spending data from github

```
hd = readr::read_csv("historical_spending.csv", col_names = TRUE)
```

```
## Rows: 13 Columns: 10
## -- Column specification ------------------------------------------------------
## Delimiter: ","
## dbl (10): Year, PercentCelebrating, PerPerson, Candy, Flowers, Jewelry, Gree...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Summary of the historical data

```
hd %>% summary()
```

```
##       Year      PercentCelebrating   PerPerson         Candy
##  Min.   :2010   Min.   :51.00      Min.   :103.0   Min.   : 8.60
##  1st Qu.:2013   1st Qu.:54.00      1st Qu.:131.0   1st Qu.:10.85
##  Median :2016   Median :55.00      Median :142.3   Median :12.70
##  Mean   :2016   Mean   :55.46      Mean   :144.4   Mean   :12.84
##  3rd Qu.:2019   3rd Qu.:58.00      3rd Qu.:162.0   3rd Qu.:14.12
##  Max.   :2022   Max.   :60.00      Max.   :196.3   Max.   :17.30
##     Flowers         Jewelry       GreetingCards      EveningOut
##  Min.   :12.33   Min.   :21.52   Min.   :5.910   Min.   :21.39
##  1st Qu.:13.49   1st Qu.:30.34   1st Qu.:7.310   1st Qu.:25.66
##  Median :14.78   Median :30.94   Median :7.870   Median :27.48
##  Mean   :14.65   Mean   :32.55   Mean   :7.676   Mean   :27.47
##  3rd Qu.:15.42   3rd Qu.:34.10   3rd Qu.:8.320   3rd Qu.:28.46
##  Max.   :16.71   Max.   :45.75   Max.   :9.010   Max.   :33.46
##     Clothing        GiftCards
##  Min.   :10.42   Min.   : 8.42
##  1st Qu.:12.00   1st Qu.:10.23
##  Median :14.04   Median :11.04
##  Mean   :14.94   Mean   :11.50
##  3rd Qu.:16.08   3rd Qu.:12.52
##  Max.   :21.46   Max.   :17.22
```
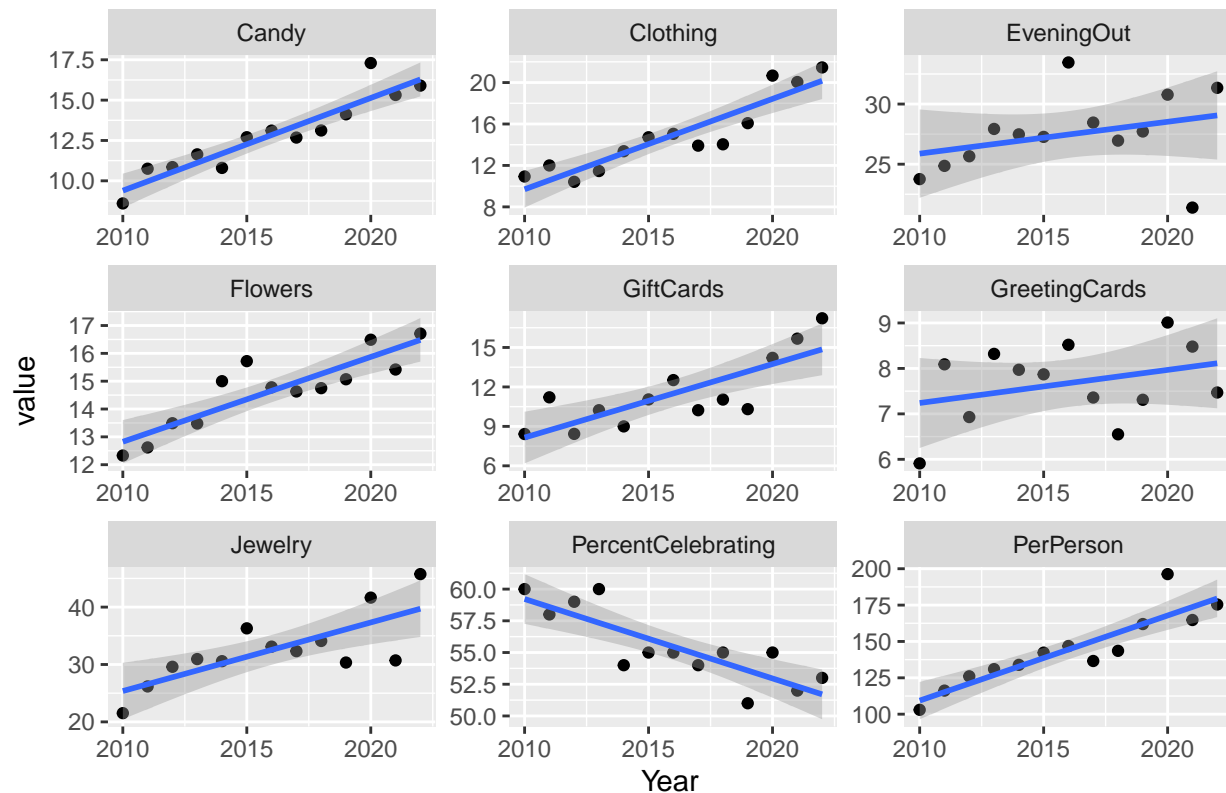
Data visualization

```
hd %>%
  pivot_longer(PercentCelebrating:GiftCards, names_to = 'param', values_to = 'value') %>%
  ggplot() +
  aes(x = Year) +
  aes(y = value) +
  geom_point() +
  stat_smooth(method='lm') +
  scale_x_continuous(breaks = c(2010, 2015, 2020)) +
  facet_wrap(~param, scales='free') +
  labs(title = "Linear regression over the historical measurements from 2010 until 2022")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```


Linear regression over the historical measurements from 2010 until 2022

```
hd %>%
  pivot_longer(PercentCelebrating:GiftCards, names_to = 'param', values_to = 'value') %>%
  ggplot() +
  aes(y = value) +
  geom_boxplot() +
  facet_wrap(~param, scales='free') +
  scale_x_continuous(breaks = c()) +
  labs(title = "Linear regression over the historical measurements from 2010 until 2022")
```

Linear regression over the historical measurements from 2010 until 2022
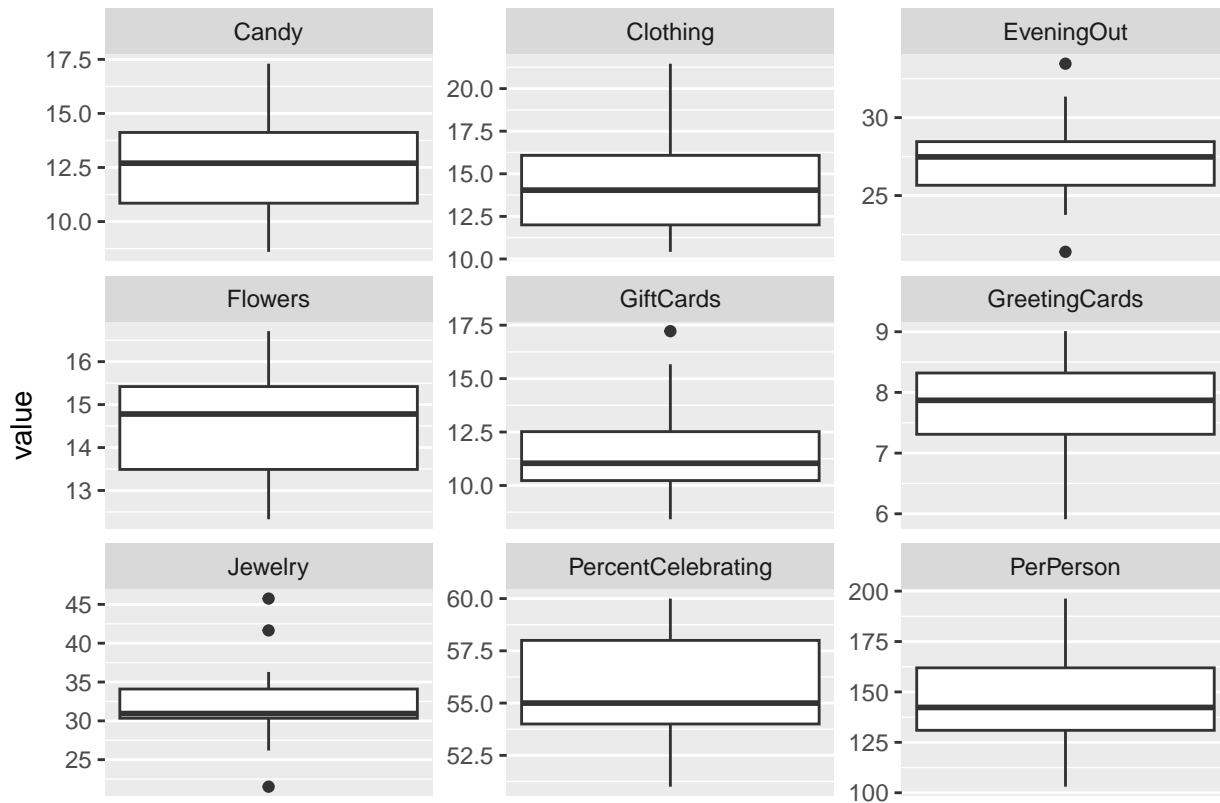
We are comparing all the categories throughout the years to see how the trend changes. We can see, that people are buying more things while the time progresses. The only category that is decreasing is the percentage of people celebrating. This an be justified by the corona pandemic. However, this trend occurred even before the pandemic started in 2020.

Testing

```
t.test(hd$GiftCards, hd$GreetingCards)
```

```
##
##  Welch Two Sample t-test
##
## data:  hd$GiftCards and hd$GreetingCards
## t = 4.8318, df = 14.426, p-value = 0.0002449
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.132875 5.520971
## sample estimates:
## mean of x mean of y
## 11.503077  7.676154
```

Using the T-test we can determine, that the percentage of people that are buying the gift cards and greeting cards is not the same. The P-value shows a significant difference in the two means. Even thou the amount of people buying the cards increases throughout the years, the percentage is not the same.