

CXC 2020 Homework four

Renee Zale

3/24/2021

The following data comes from The Customer Experience Code (CXC) (<https://cspace.com/customerexperienced/>), an annual benchmarking study conducted by C Space that evaluates the performance of 183 popular brands across 18 industries. The data comes from a survey fielded to 13,926 consumers, who were each asked to evaluate two brands against a series of proprietary metrics designed to measure how well brands connect with their customers. Past C Space research has proven that these measures are key drivers of growth for brands. Respondents also evaluated the brands on Net Promoter Score (NPS) and Earned Advocacy Score (<https://hbr.org/2019/10/where-net-promoter-score-goes-wrong#:~:text=By%20taking%20the%20percentage%20of,%2C%20and%20more%2Dactionable%20data.>), a proprietary metric developed by C Space as an alternative to NPS.

Initially, I wanted to see how mean NPS scores vary across industries.

As a first step to my analysis, I created a new dataset called meanNPS, where I summarized the average NPS rating (10 point scale) by industry.

```
meanNPS <- CXC2020 %>%
  group_by(Industry) %>%
  summarise(NPS = mean(NPS))

print(meanNPS)
```

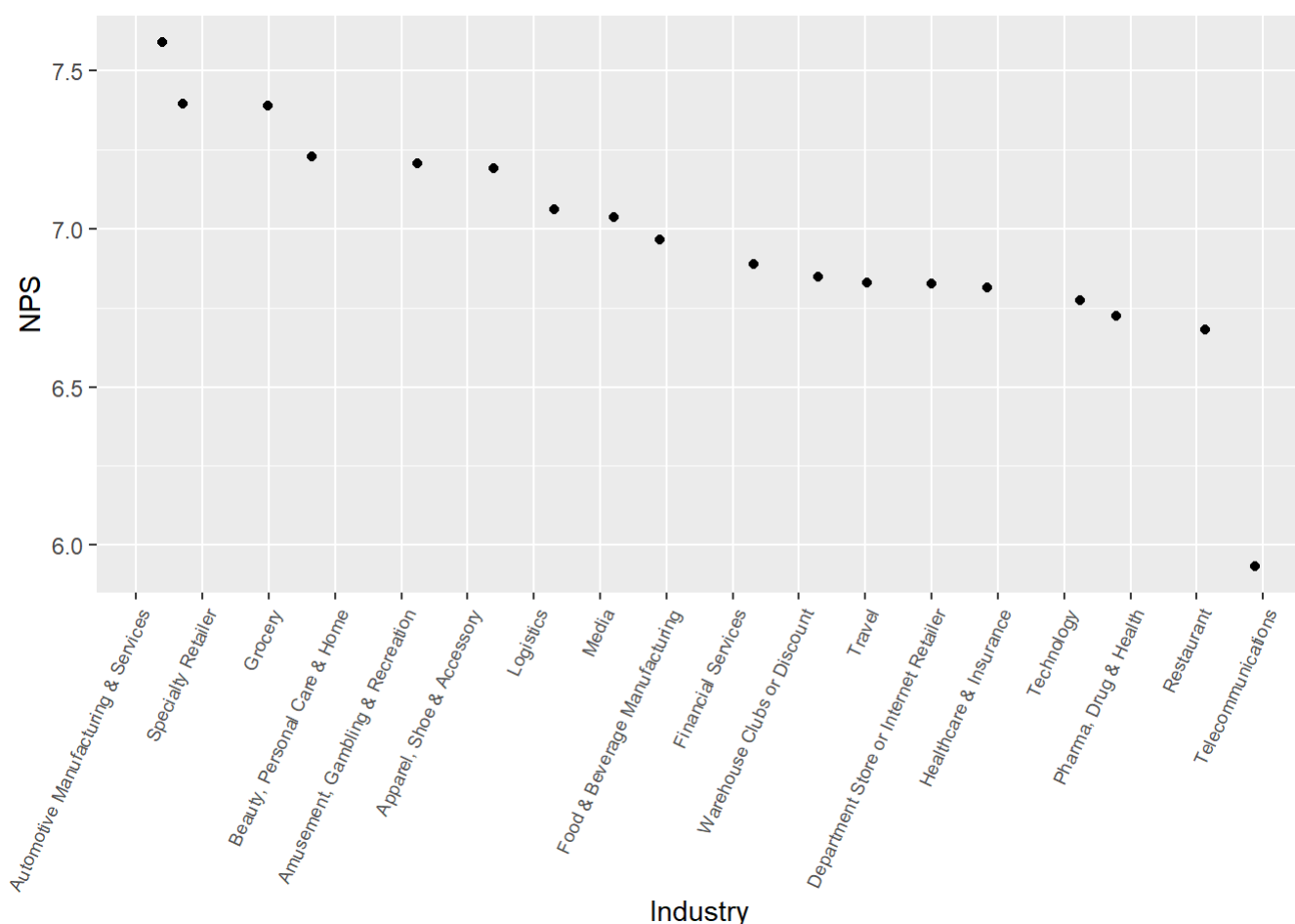
```
## # A tibble: 18 x 2
##   Industry          NPS
##   <chr>          <dbl>
## 1 Amusement, Gambling & Recreation    7.21
## 2 Apparel, Shoe & Accessory           7.19
## 3 Automotive Manufacturing & Services  7.59
## 4 Beauty, Personal Care & Home        7.23
## 5 Department Store or Internet Retailer 6.83
## 6 Financial Services                  6.89
## 7 Food & Beverage Manufacturing       6.97
## 8 Grocery                           7.39
## 9 Healthcare & Insurance              6.81
## 10 Logistics                         7.06
## 11 Media                             7.04
## 12 Pharma, Drug & Health              6.73
## 13 Restaurant                        6.68
## 14 Specialty Retailer                 7.40
## 15 Technology                        6.78
## 16 Telecommunications                5.93
## 17 Travel                           6.83
## 18 Warehouse Clubs or Discount        6.85
```

In my chart I wanted my Industry categories to run along the X axis in ascending order. I created a new dataset called NPSmeansDesc and sorted industry by NPS so that this will appear in my chart. *Note: I tried many different ways to do this but found the only way that worked for me was by creating a new dataset. I'd love to know if there is a different way to do this within the chart code itself.*

```
NPSmeansDesc <- meanNPS
NPSmeansDesc$Industry <- factor(NPSmeansDesc$Industry,
                                levels = NPSmeansDesc$Industry[order(NPSmeansDesc$NPS, decreasing = T
                                RUE)])
```

Then, I plotted the sorted average NPS scores by industry.

```
ggplot(data = NPSmeansDesc) +
  geom_point(mapping = aes(x = Industry, y = NPS), position = "jitter") +
  theme(axis.text.x = element_text(size = 7, angle = 65, hjust = 1))
```

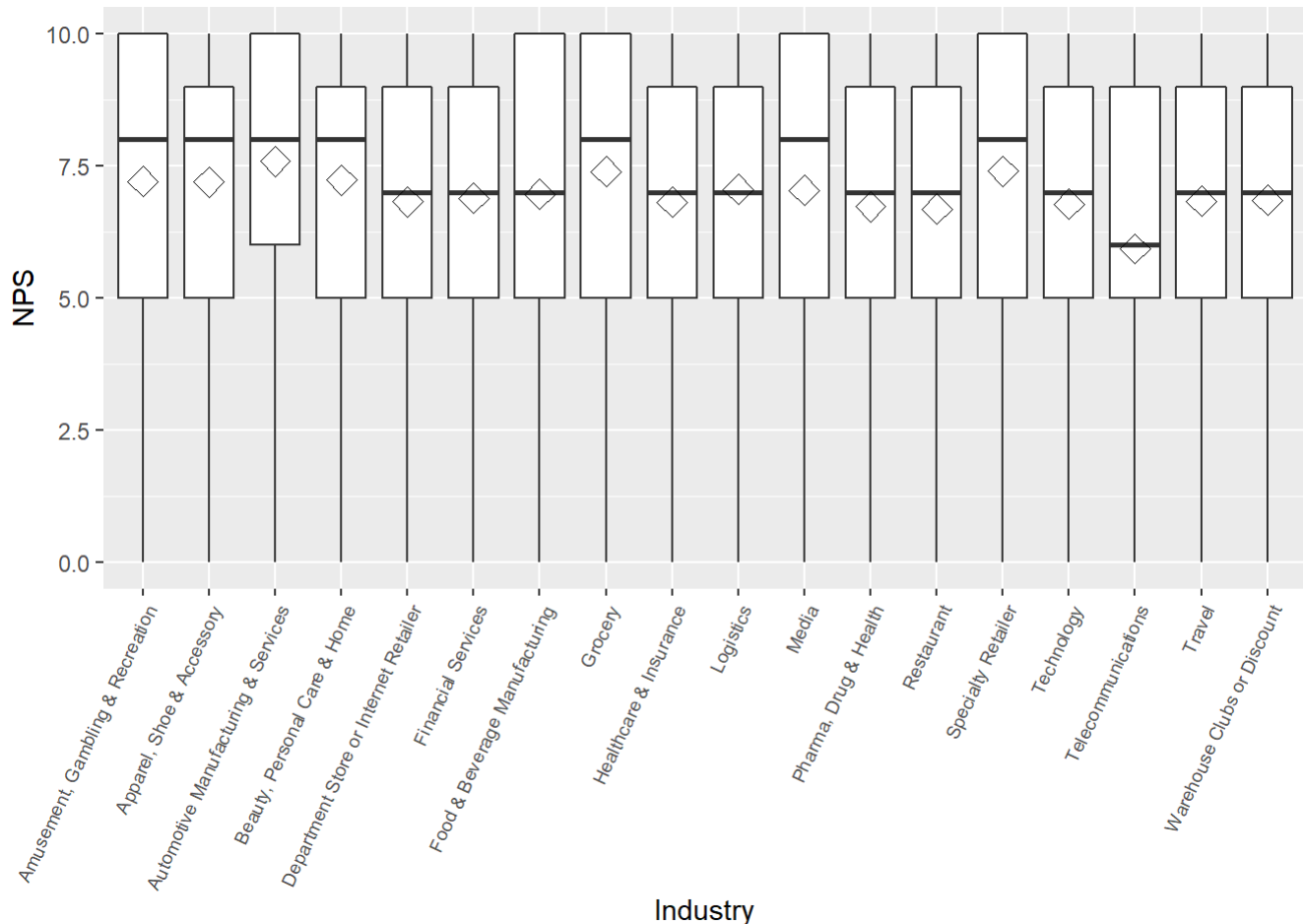


From this chart we can see that consumers are most likely to recommend brands in the Auto Manufacturing & Services, Specialty Retailer, Grocery, and Beauty, Personal Care & Home industries. Consumers are least likely to recommend brands in the telecommunications space.

I was curious about the distribution of NPS ratings by industry, so I created a set of boxplots.

```
ggplot(CXC2020, aes(x=Industry, y=NPS)) +
  geom_boxplot(outlier.colour="red", outlier.shape=8, outlier.size=4) +
  theme(axis.text.x = element_text(size = 7, angle = 65, hjust = 1)) +
  stat_summary(fun.y=mean, geom="point", shape=23, size=4)
```

```
## Warning: Ignoring unknown parameters: outlier.colour
```



Across almost all industries, there is no variation in the lower scores (the lower quartile). Where we do see variation is in the medians, means, and the first quartiles. A few points that jump out to me as interesting: 1. The mean NPS rating for several industries are lower than the median (i.e. Amusement, Apparel, Automotive, Beauty, Grocery, Media, and Specialty). This could mean that these industries are more polarizing, with some very low scores bringing down the average. 2. Several industries do not have an 'upper whisker' (like Amusement, Automotive, and Food & Beverage). This means that their distribution of scores skews very high, with many respondents saying they would be likely to promote brands in these industries. 3. Telecommunications has the lowest median and mean, with a quarter of the respondents rating these brands as a 5-6 out of 10.

After creating the boxplot, I then wanted to visualize the top 10 performing brands based on average NPS rating. First, I grouped the mean NPS scores by brand.

```
meanNPSbrand <- CXC2020 %>%
  group_by(Company) %>%
  summarise(NPS = mean(NPS)) %>%

print(meanNPSbrand)
```

```
## # A tibble: 182 x 2
##   Company      NPS
##   <chr>      <dbl>
## 1 3M          7.10
## 2 Adidas     6.95
## 3 Adobe      6.46
## 4 Aetna      6.47
## 5 Airbnb     7.16
## 6 Alaska Airlines 7.03
## 7 Aldi       7.71
## 8 Amazon     7.90
## 9 Amazon Fresh 7.73
## 10 American Airlines 5.94
## # ... with 172 more rows
```

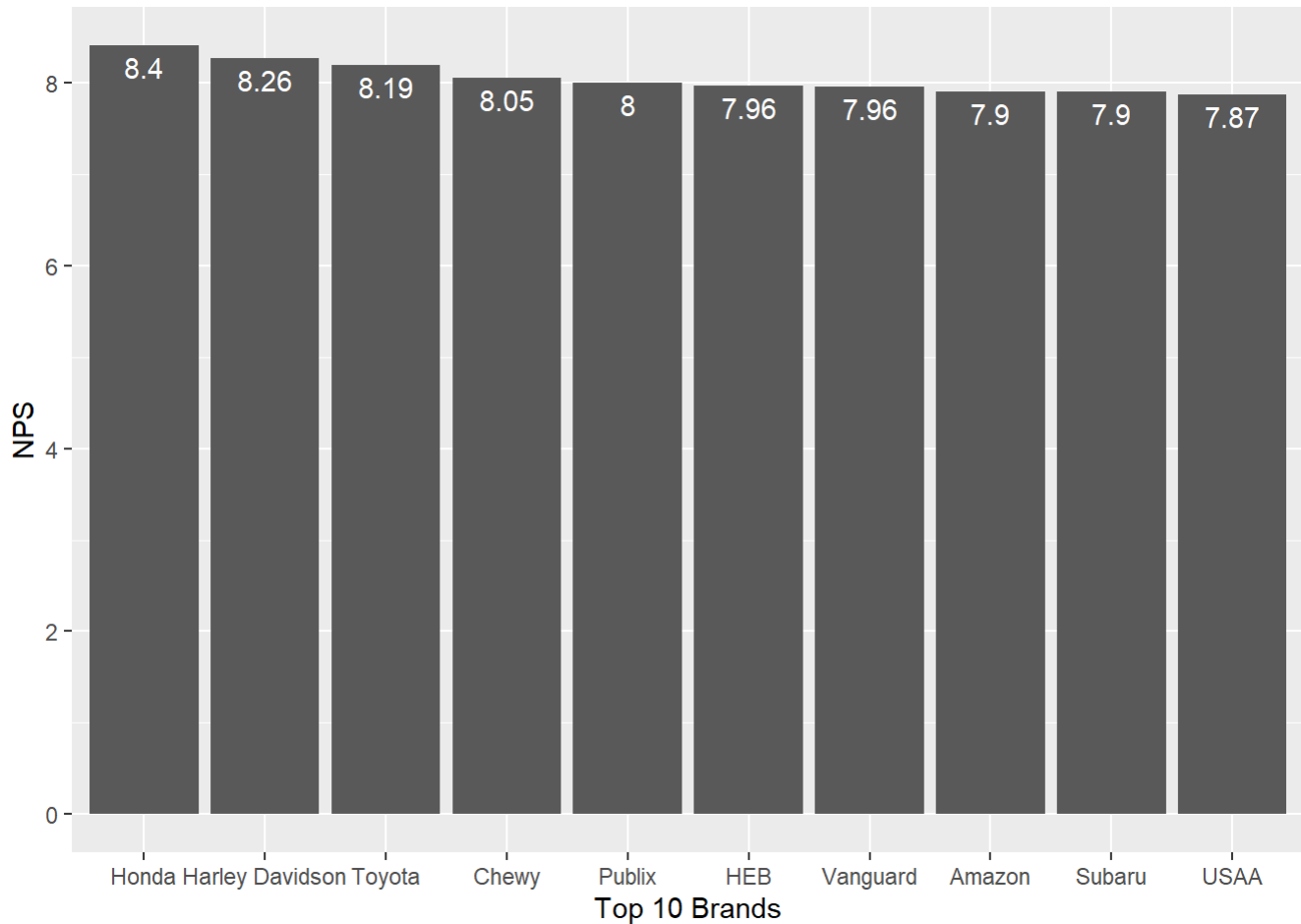
Then, I created a new dataset to sort these mean brand NPS ratings in descending order and select only the top 10 before plotting. *This is another example of an instance where I wonder if it is possible to do this within the original dataset, or if it is necessary to create a new set as I have done.*

```
NPSbrandDesc <- meanNPSbrand %>%
  arrange(desc(NPS)) %>%
  slice(1:10)

print(NPSbrandDesc)
```

```
## # A tibble: 10 x 2
##   Company      NPS
##   <chr>      <dbl>
## 1 Honda      8.40
## 2 Harley Davidson 8.26
## 3 Toyota     8.19
## 4 Chewy      8.05
## 5 Publix      8
## 6 HEB        7.96
## 7 Vanguard   7.96
## 8 Amazon     7.90
## 9 Subaru     7.90
## 10 USAA      7.87
```

```
ggplot(NPSbrandDesc, aes(x=reorder(Company, -NPS), y=NPS))+
  geom_bar(stat='identity') +
  geom_text(aes(label = round(NPS, 2)), vjust = 1.5, colour = "white")+
  xlab("Top 10 Brands")
```



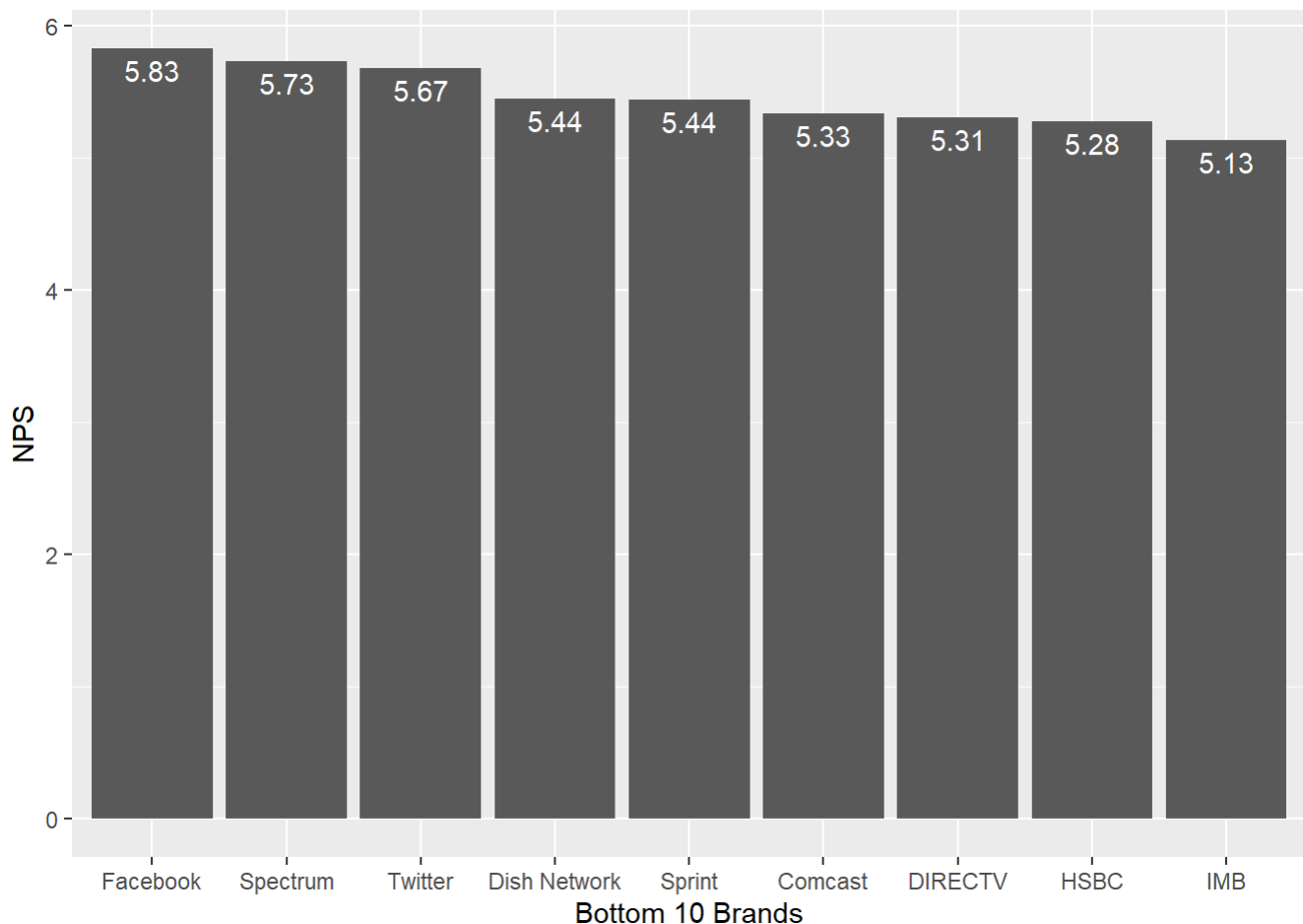
Through this visualization we can see that the Honda, Harley Davidson, Toyota and Chewy are among the most recommended brands. I would be interested to see if this has changed from last year's wave of data, given that consumers might potentially feel more positive about retailers like Publix, HEB, Chewy, and Amazon due to the pandemic's impact on shopping.

Next, I wanted to see the 10 worst performing brands of the bunch. I repeated the steps above, but this time sliced to only include the last 10 brands in descending order.

```
worstNPSbrandDesc <- meanNPSbrand %>%  
  arrange(desc(NPS)) %>%  
  slice(174:183)  
  
print(worstNPSbrandDesc)
```

```
## # A tibble: 9 x 2
##   Company      NPS
##   <chr>      <dbl>
## 1 Facebook    5.83
## 2 Spectrum    5.73
## 3 Twitter     5.67
## 4 Dish Network 5.44
## 5 Sprint      5.44
## 6 Comcast     5.33
## 7 DIRECTV     5.31
## 8 HSBC        5.28
## 9 IMB         5.13
```

```
ggplot(worstNPSbrandDesc, aes(x=reorder(Company, -NPS), y=NPS))+
  geom_bar(stat='identity') +
  geom_text(aes(label = round(NPS, 2)), vjust = 1.5, colour = "white") +
  xlab("Bottom 10 Brands")
```



Looking at the worst average NPS ratings, it is interesting to see that the bulk of them are cable utilities like Comcast or DIRECTV. Facebook and Twitter also appear in the worst performers, perhaps a reflection of how these companies handled information related to the pandemic and 2020 election.

One visualization I wanted to try to make would be calculating the Net Promoter Score for each brand and plotting it by industry. The Net Promoter Score is calculated by subtracting the percent of detractors (rated the brand 0-6) from the percent of promoters (rated the brand 9-10). I would then want to sort the industries in order of NPS score

in a chart. I haven't quite figured out how to do this regrouping and calculation to create the Net Promoter Score variable.