Exploring Data

Ben Goldstone

02/06/23

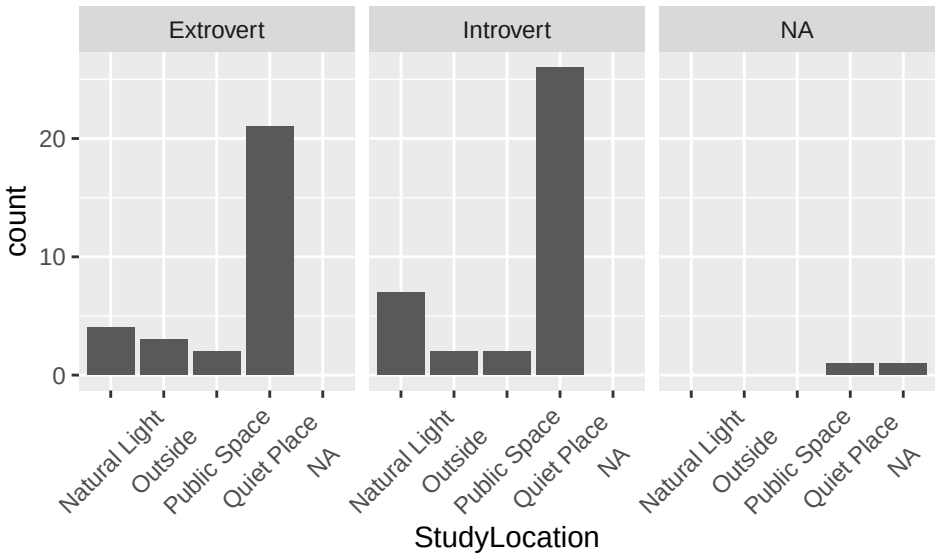
Importing Dataset

```
library(readr)
Spring2023Dataset <- read_csv("~/CSVs/Spring2023Dataset.csv",
    col_types = cols(SpendOnYou = col_integer(),
        TextsSent = col_integer(), TextsReceived = col_integer(),
        GroupText = col_integer(), HighwayAlone = col_integer(),
        HighwayPassenger = col_integer()))
head(Spring2023Dataset)
```

```
## # A tibble: 6 x 18
##   SpendOnYou LongestRun PizzaToppings ToppingPref GiveUp TextsSent TextsReceived
##        <int>      <dbl>         <dbl> <chr>       <chr>      <int>         <int>
## 1        400        1               3 Mostly veg~ Meat          50            50
## 2        850        7               4 Mostly veg~ Caffe~        20            25
## 3        550        2.5             4 Mostly mea~ <NA>          23            24
## 4       1000        0.5             2 Mostly veg~ Meat          20            20
## 5        400       10               2 Mostly mea~ Desse~        15            20
## 6        300        1               2 Mostly mea~ Caffe~         0             1
## # ... with 11 more variables: GroupText <int>, TextOften <chr>, Active <chr>,
## #   Generous <chr>, Gender <chr>, HighwayAlone <int>, HighwayPassenger <int>,
## #   CoffeeFreq <chr>, StudyLocationOriginal <chr>, Personality <chr>,
## #   StudyLocation <chr>
```

Study Location

```
gf_bar( ~ StudyLocation | Personality,
        data = Spring2023Dataset) + theme(axis.text.x = element_text(
          angle = 45,
          hjust = 0.5,
          vjust = 0.5
        ))
```



Study Location Variables

Response Variable: Study Location (Categorical) - locations students prefer to study in.

Explanatory Variable: Personality(Categorical) - personality type of students. ## Observations Based on the data set, it seems that people who are introverted or extroverted, prefer studying in a quiet place

```
tally(StudyLocation~Personality, data=Spring2023Dataset, format='prop')
```

```
##                Personality
## StudyLocation    Extrovert  Introvert       <NA>
##   Natural Light 0.13333333 0.18918919 0.00000000
##   Outside       0.10000000 0.05405405 0.00000000
##   Public Space  0.06666667 0.05405405 0.00000000
##   Quiet Place   0.70000000 0.70270270 0.50000000
##   <NA>          0.00000000 0.00000000 0.50000000
```

Conclusions

In terms of this data set, I would expect more extroverts to work in a public space. According to this data set though, it seems extroverts prefer a quiet place around 56.7% higher than the proportion of extroverts that prefer to study in a public space.

Generosity

```
gf_boxplot(SpendOnYou ~ Generous, data = Spring2023Dataset, fill = "darkgreen") +
  scale_x_discrete(
    labels = c(
      "No, I could be a bit more generous" = "Not very generous person",
      "Yes, I'm a pretty generous person" = "Generous person"
    )
  )
```

```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```

Generosity Variables

Response Variable: SpendOnYou - how much money students would spend on themselves if they were awarded $1000.

Explanatory Variable: Generous - how generous is the student.

Observations

According to the box plot, it looks like the middle 50% of people who would spend more, also tend to be less generous.

```
favstats(SpendOnYou~Generous, data=Spring2023Dataset)
```

```
##                             Generous min  Q1 median  Q3  max     mean       sd
## 1 No, I could be a bit more generous 200 500    700 900 1000 681.8182 259.1069
## 2  Yes, I'm a pretty generous person   0 300    500 700 1000 509.6316 282.8810
##    n missing
## 1 11       0
## 2 57       0
```

Conclusions

When looking at the 5 number summary, it showed me that 5x the amount of people marked that they were generous, compared to the people who said they could be more generous. Given that, it is difficult to conclude if the data set is accurate given the higher data from people that are generous.