

Survey Results

Code ▾

Evaluating Survey Results

Setup

Hide

```
library(ggplot2)
library(tidyverse)
# library(summarytools)
```

Loading the table

Hide

```
# file = r"(C:\Users\Valen\OneDrive\Dokumente\MASTER\WiSe_25_26\AppliedStatistics\Project\Week1\Week1_Survey_Results.csv)"

file = r"(C:\Users\Valentin\Documents\GIT_REPS\TUHH\Applied-Statistics\week01\Week1_Survey_Results.csv)"

survey <- read.csv(file, header = TRUE, sep = ",")
```

Clean the Data

Hide

```
# select all coffee drinkers and remove trailing and leading whitespaces
filtered_survey <- survey |> filter(Drinks.Coffee == 'Y') |> mutate(across(where(is.character), str_trim))

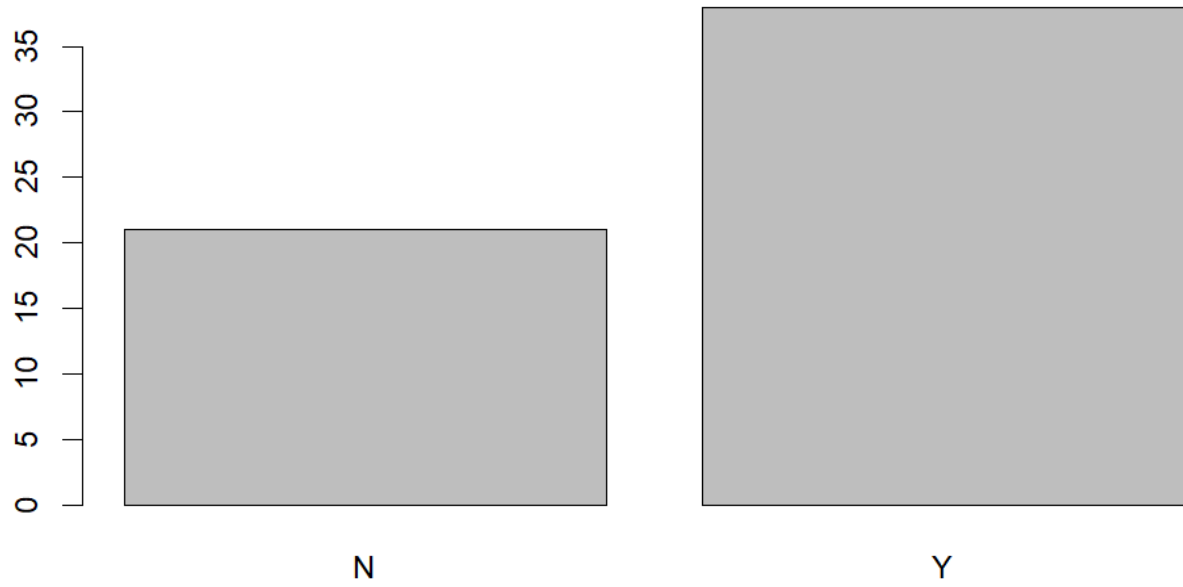
all_semesters <- survey$Semester
semesters <- filtered_survey$Semester
locations <- filtered_survey$Location
ratings <- filtered_survey$Rating
types <- filtered_survey$Type
reasons <- filtered_survey$Reason
```

View the Data

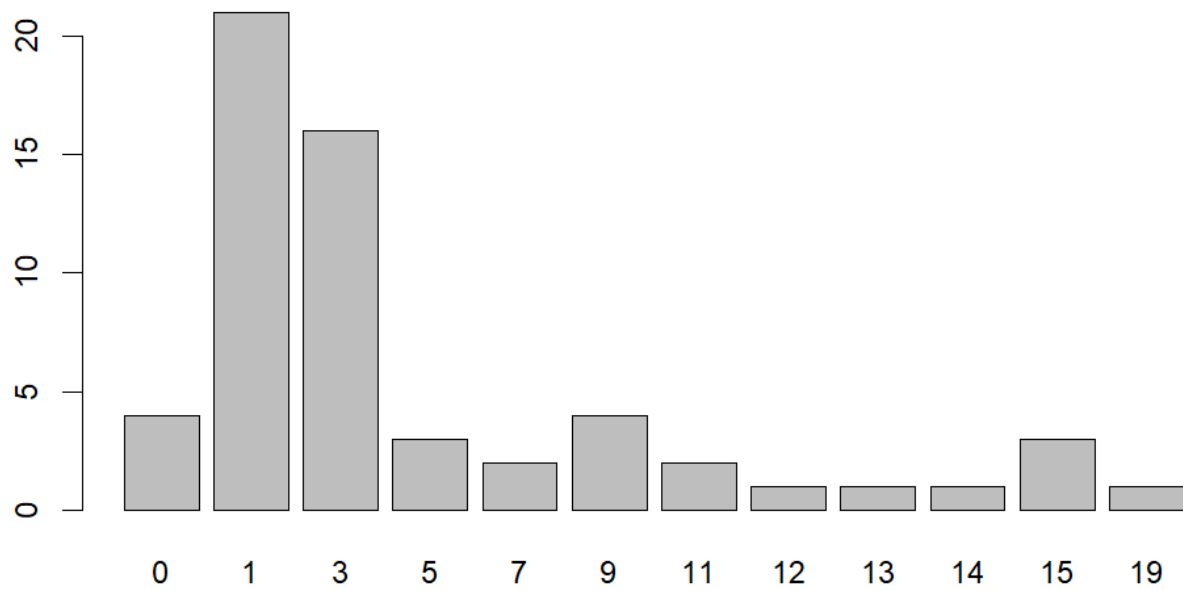
histograms

Hide

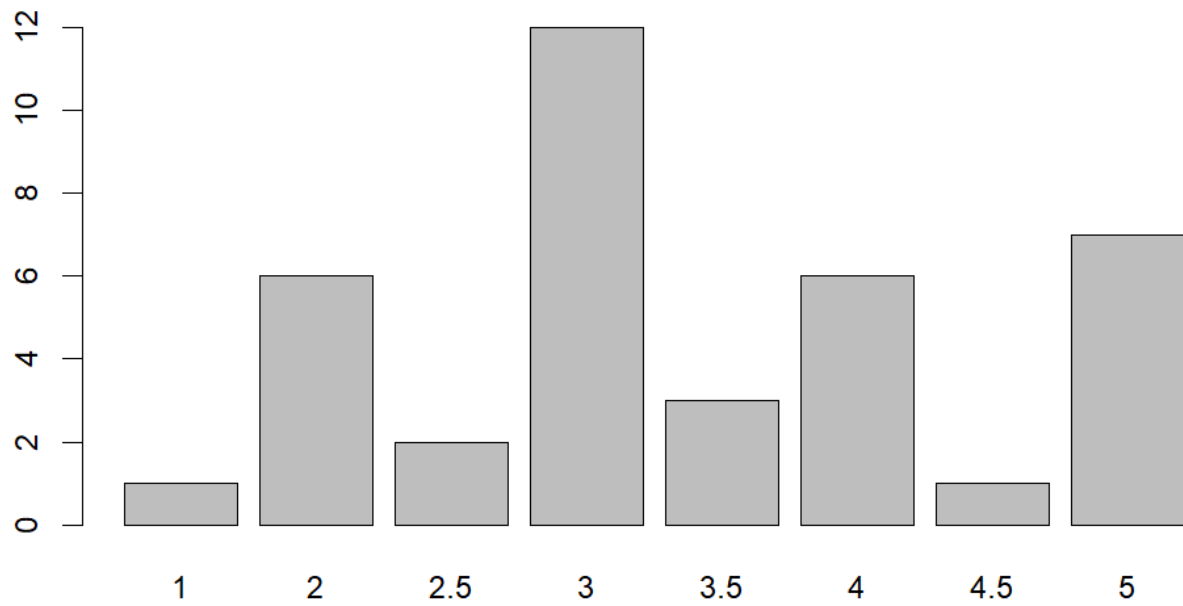
```
barplot(table(survey$Drinks.Coffee))
```

[Hide](#)

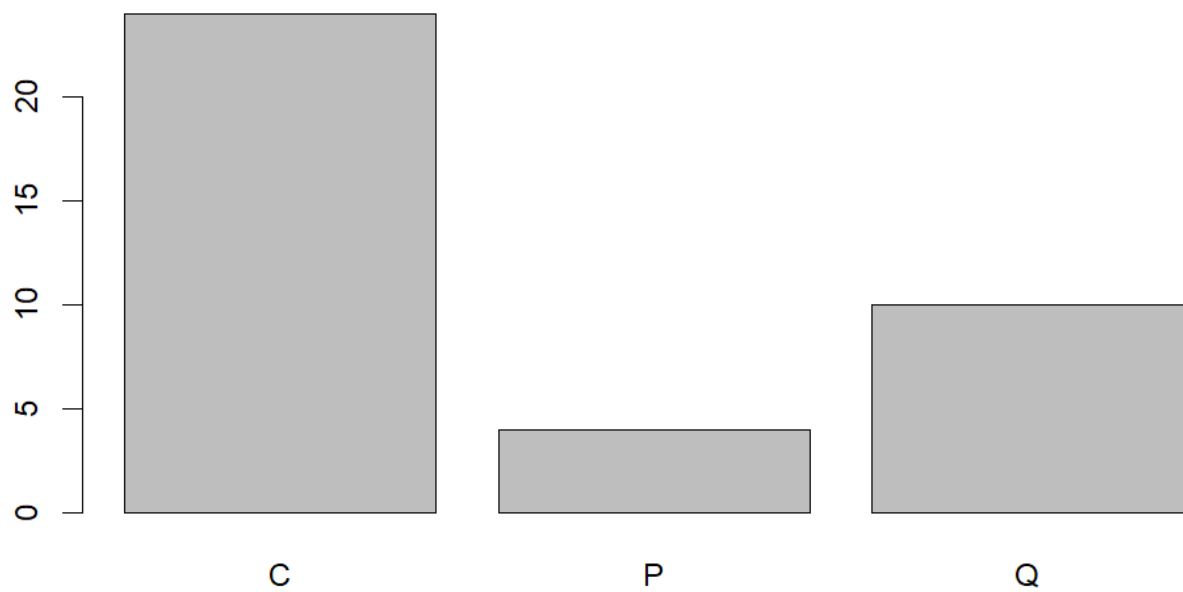
```
# hist(survey$Semester, breaks = seq(0, 20, 1))  
# hist(survey$Rating, breaks = seq(0, 5, 1))  
  
barplot(table(all_semesters))
```

[Hide](#)

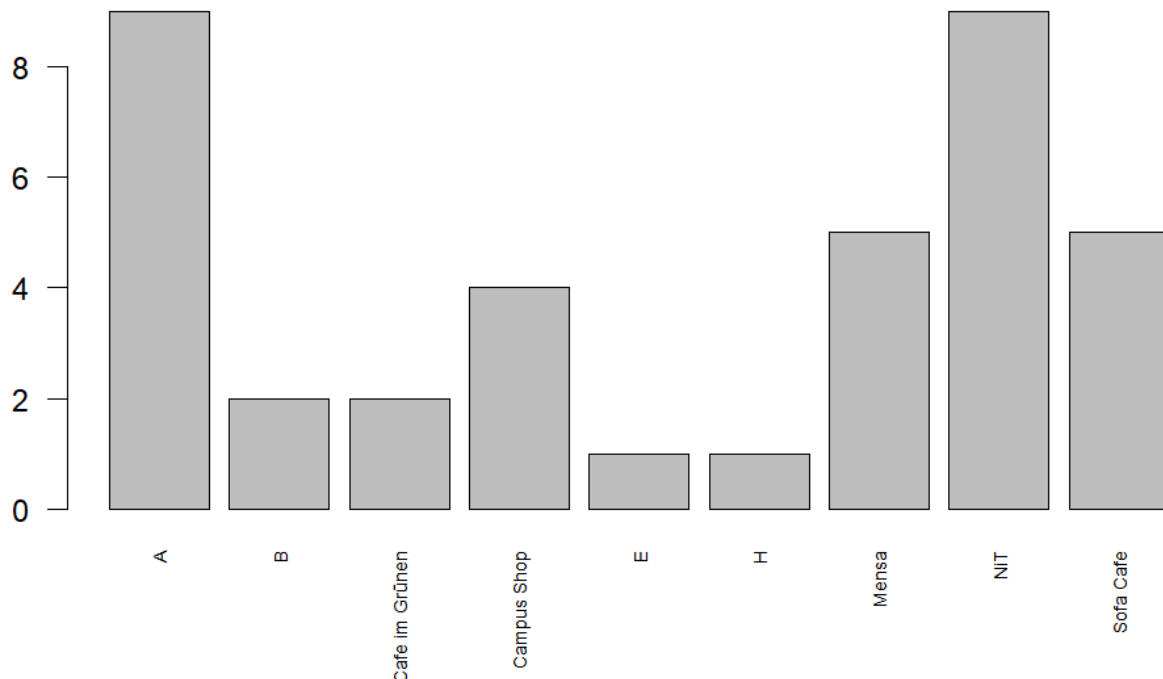
```
barplot(table(ratings))
```

[Hide](#)

```
barplot(table(reasons))
```

[Hide](#)

```
barplot(table(locations), cex.names = 0.6, las = 2)
```



Analyze the data

How are Locations rated?

[Hide](#)

```
# descr(filtered_survey)

location_rating <- filtered_survey |>
  group_by(Location) |>
  summarize(mean_rating = mean(Rating), N = n()) |>
  arrange(desc(mean_rating))

print(location_rating)
```

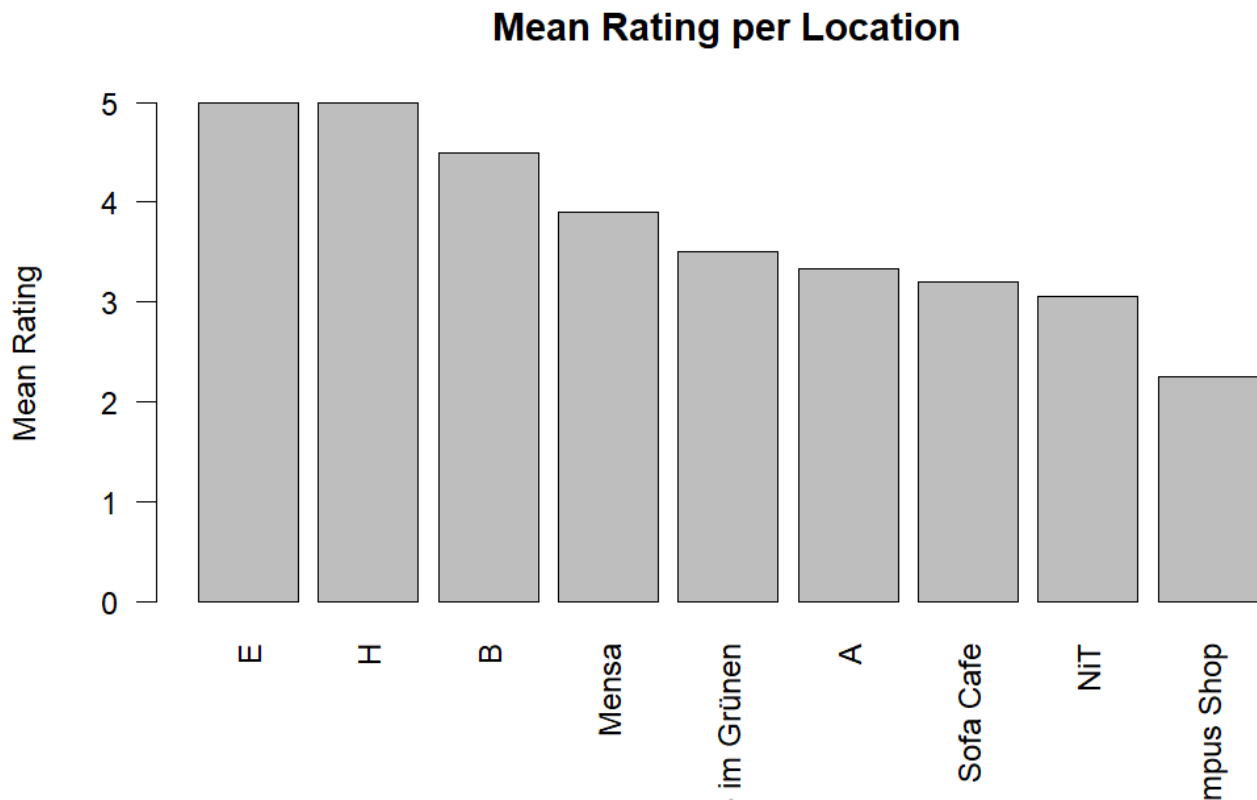
Location <chr>	mean_rating <dbl>	N <int>
E	5.000000	1
H	5.000000	1
B	4.500000	2
Mensa	3.900000	5
Cafe im Grünen	3.500000	2
A	3.333333	9

Location <chr>	mean_rating <dbl>	N <int>
Sofa Cafe	3.200000	5
NiT	3.055556	9
Campus Shop	2.250000	4

9 rows

Hide

```
barplot(location_rating$mean_rating,
        names.arg = location_rating$Location,
        ylab = "Mean Rating",
        main = "Mean Rating per Location",
        las = 2)
```



How do the semesters correspond to the Students Ratings

Hide

```
semester_rating <- filtered_survey |>
  group_by(Semester) |>
  summarize(mean_rating = mean(Rating), N = n(), .groups = "drop")

print(semester_rating)
```

Semester <int>	mean_rating <dbl>	N <int>
0	3.000000	4
1	3.458333	12
3	3.321429	14
5	3.500000	2
11	5.000000	1
13	2.000000	1
15	3.000000	3
19	5.000000	1

8 rows

[Hide](#)

```
ggplot(semester_rating, aes(x = Semester, y = mean_rating)) +
  geom_point(fill = "skyblue") +
  geom_smooth(method = "lm", color = "red") +
  labs(title = "Mean Rating per Semester with linear trend", y = "Mean Rating", x = "Semester") +
  theme_minimal()
```

Mean Rating per Semester with linear trend

