

### Loading a Package

```
library(PACKAGE NAME)
```

### Reading in Data

```
NAME OF DATASET <- read_csv("PATH & NAME OF DATASET.csv")
```

**Note:** The name of the dataset will change, but it will always need to have the .csv at the end of its name!

**Note:** Do not put spaces in the name you give the data set.

### Preview a Dataset

```
glimpse(NAME OF DATASET)
```

```
head(NAME OF DATASET) – shows first 6 rows
```

```
names(NAME OF DATASET) – outputs the names of the columns/variables
```

### Plotting a One Categorical Variable Bar Plot with Counts

```
ggplot(data = NAME OF DATASET,  
  mapping = aes(x = NAME OF VARIABLE)) +  
  geom_bar(stat = "count") +  
  labs(title = "TITLE FOR GRAPH",  
    x = "TITLE FOR THE X-AXIS",  
    y = "TITLE FOR THE Y-AXIS")
```

**Note:** This bar plot has the variable names on the x-axis. If the names are squished, then you should use **y = NAME OF VARIABLE** instead of **x = NAME OF VARIABLE**.

### Plotting a One Categorical Variable Bar Plot with Proportions

```
ggplot(data = NAME OF DATASET,  
  mapping = aes(x = NAME OF VARIABLE)) +  
  geom_bar(stat = "count", aes(y = ..prop.., group = 1)) +  
  labs(title = "TITLE FOR GRAPH",  
    x = "TITLE FOR THE X-AXIS",  
    y = "TITLE FOR THE Y-AXIS")
```

**Note:** This bar plot has the variable names on the x-axis. If the names are squished, then you should use **y = NAME OF VARIABLE** instead of **x = NAME OF VARIABLE**.

### Creating a Summary Table of Observations of One Categorical Variable

```
NAME OF DATASET |>  
count(NAME OF VARIABLE)
```

### Conducting an Exact Binomial Hypothesis Test for One Proportion

```
binom.test(x = NUMBER OF SUCCESSES, n = SAMPLE SIZE, p = NULL VALUE, alternative = "DIRECTION")
```

**Note:** The alternative direction can be "greater", "less", or "two.sided"

### Performing a Chi-Squared Goodness-of-Fit Test (One Categorical Variable)

```
chisq_test(x = NAME OF DATASET,  
  response = NAME OF VARIABLE,  
  p = c("CATEGORY 1" = EXPECTED PROP 1,  
        "CATEGORY 2" = EXPECTED PROP 2,  
        "CATEGORY LAST" = EXPECTED PROP LAST  
  )  
)
```

\*Make sure to check conditions first!

### Plotting a Two Categorical Variable Bar Plot

```
ggplot(data = NAME OF DATASET,  
  mapping = aes(x = EXPLANATORY VARIABLE,  
                fill = RESPONSE VARIABLE)  
  ) +  
  geom_bar(stat = "count",  
           position = "fill") +  
  labs(title = "TITLE FOR GRAPH",  
        x = "TITLE FOR THE X-AXIS",  
        y = "TITLE FOR THE Y-AXIS")
```

**Note:** If you want a side-by-side bar plot you need to change position to "dodge". If you want a stacked bar plot, you need change position to "stack".

### Creating a Summary Table of Observations from Two Categorical Variables

```
NAME OF DATASET |>  
count(NAME OF VARIABLE 1, NAME OF VARIABLE 2)
```

### Creating a Contingency Table of Observed Counts from Two Categorical Variables

```
NAME OF DATASET |>  
  count(EXPLANATORY VARIABLE, RESPONSE VARIABLE) |>  
  pivot_wider(names_from = RESPONSE VARIABLE,  
               values_from = n) |>  
  adorn_totals(where = c("row", "col"))
```

**Note:** Your explanatory variable should be in the rows and your response variable should be in the columns. So, the variable you insert into names\_from should be the response variable you are interested in.

### Creating a Contingency Table of Observed Proportions from Two Categorical Variables

```
NAME OF DATASET |>  
  count(EXPLANATORY VARIABLE, RESPONSE VARIABLE) |>  
  pivot_wider(names_from = RESPONSE VARIABLE,  
               values_from = n) |>  
  adorn_totals(where = c("row", "col")) |>  
  adorn_percentages(denominator = "row")
```

**Note:** Since your explanatory variable (groups) should be in your rows from above, we want to calculate our proportions in respect to the group totals.

### Performing a Chi-Square Test (Two Categorical Variables)

```
chisq_test(x = NAME OF DATASET,  
           response = RESPONSE VARIABLE,  
           explanatory = EXPLANATORY VARIABLE)
```

\*Make sure to check conditions first!