R Cheatsheet

STAT218

|  |
| --- |
| Note |
| This document is a work in progress and will be updated prior to each test. |

### Basics

Loading data:

* [.RData file] load('<FILEPATH>')
* [from package] data(<DATASET NAME>, package = '<PACKAGE NAME>')
* [reading a CSV file] read.csv('<FILEPATH>')

Viewing dataframes:

* [preview] head(<NAME>)
* [data viewer] view(<NAME>)
* [check structure] str(<NAME>)

Extracting a variable from a dataframe:

* DATAFRAME$VARIABLE

### Summary statistics

If x is a vector of values of a numeric variable…

* mean(x) computes the average
* median(x) computes the median
* min(x) and max(x) compute the minimum and maximum
* quantile(x, probs = <PERCENTILE>) computes the percentile
* summary(x) computes the five-number summary, plus the mean
* range(x) computes the range (min, max)
* IQR(x) computes the interquartile range
* var(x) computes the variance
* sd(x) computes the standard deviation

If df is a dataframe with a numeric variable y and a categorical variable x…

* df |> group\_by(x) |> summarize(<OUTPUT.NAME> = <FUNCTION>(y)) computes the statistic specified by <FUNCTION> separately for each category of the variable x (requires tidyverse package)

*See especially* [*Lab 2: Descriptive statistics*](lab2-descriptive.qmd).

### Tables

If x and y are a vectors of values of two categorical variables…

* table(x) computes the frequency distribution (counts)
* table(x) |> proportions() computes the frequency distribution (proportions)
* table(x, y) computes a contingency table
* table(x, y) |> proportions(margin = NULL) computes proportions using grand total
* table(x, y) |> proportions(margin = 1) computes proportions using row total (group by x)
* table(x, y) |> proportions(margin = 2) computes proportions using column total (group by y)

*See especially* [*Lab 1: R basics*](lab1-rbasics.qmd) *and* [*Lab 3: Bivariate summaries*](lab3-bivariate.qmd).

### Graphics

If x and y are vectors of values of two numeric variables…

* hist(x, breaks = <NUMBER OF BINS>) generates a histogram
* boxplot(x) generates a boxplot
* plot(x, y) generates a scatterplot

If x and y are vectors of values of two categorical variables…

* table(x) |> barplot() generates a bar plot (counts)
* table(x) |> proportions() |> barplot() generates a bar plot (proportions)
* table(x, y) |> proportions(margin = 2) |> barplot(legend = T) generates a stacked bar plot grouped by y
* table(y, x) |> proportions(margin = 2) |> barplot(legend = T) generates a stacked bar plot grouped by x

If x is a vector of values of a categorical variable and y is a vector of values of a numeric variable…

* boxplot(y ~ x) generates a boxplot with x on the x axis (vertical orientation)
* boxplot(y ~ x, horizontal = T) generates a boxplot with y on the x axis (horizontal orientation)

*See especially* [*Lab 2: Descriptive statistics*](lab2-descriptive.qmd) *and* [*Lab 3: Bivariate summaries*](lab3-bivariate.qmd)*.*