

Using RStudio and Rmarkdown

Install the tidyverse

First step is to get the package we will be working with, the tidyverse. We can run `install.packages("tidyverse")` in the console, or go to the packages pane and click “install” to install it.

Then load it up!

Upload the data file we want to work from

In the bottom-right panel, we see all the files associated with this project. Right now, there aren’t many! We need to upload the data file so that we can work with it.

Load the data into R

Our data now lives in the cloud, but R is still not ready to work with it. Notice, the environment is empty. We can load the data in using “Import Dataset” button. Let’s give it the name “scr” for short. Now it’s in the environment and ready to be used!

The data is in a format called a data frame, which is a fancy way of saying it is arranged with each variable in a column. We can inspect it by clicking on it in the environment.

We can use code chunks to embed analyses

If we click “Insert Chunk” then we can embed R code into our document like so, and run it!

We can make basic plots

We can use ggplot to make a simple scatterplot of screen time vs happiness.

Or we could make a line graph.

We can make lots of different types of plots

Let’s look at another type of data: group comparisons. In data science, we sometimes call this an A/B test. We’ll load in a new dataset where we can compare fear of flying vs fear of driving.

We’ll make boxplots of the two groups’ data.

Compare the means using a t-test

We can do a statistical test with a single line of code.

Confidence interval plots

We can summarize our findings using a confidence interval plot. First we need to obtain the two relevant quantities: means and confidence intervals for each travel mode.

In R we can do this using the `group_by()` and `summarize()` functions. We will use the pipe operator `%>%` to apply these functions in turn.

Let's make a plot with the estimates and confidence intervals. We'll use our new summary table, called "estimates."

Let's do a basic regression analysis

We've seen how to make a scatterplot. Now let's do a formal regression analysis of some X/Y data.

We'll use a dataset that measured people's time spent on social media as well as their level of depression and social media addiction. Let's make some basic descriptive plots.

Hypothesis time!

We might hypothesize that those who spend more time on social media may have higher levels of depression.

We'll make a scatter plot showing time spent on the x-axis and depression score on the y-axis.

Fit the linear model

We'll fit a linear regression model to these data with depression as the outcome and time as the predictor.

We'll also add the estimated regression line to our plot.