```
# DESCRIPTIVE TITLE FOR SCRIPT
# FIRST_NAME LAST_NAME
# Date in YYYY-MM-DD format

##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
## 0. Intro ----
##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

# This assignment tests that you can:
# - read in data with `read_csv()`
# - create frequency tables with `janitor::tabyl()` and save those with `write_csv()`
# - create simple plots with `esquisse::esquisser()` and save those plots
# - work well in a group

# Your final grade will be the average of all the work done in your pod.
# So if you finish ahead of time, try to help out your pod members!
# If a pod member is missing from class, you can ignore that script. No need to cover fo
r them.
# You have about 1 hour to complete the task.
# But you can continue to work on it later on. The final deadline is next Tuesday, Oct 2
5, 23:59 pm CET.
# See section 5 below for details on submission

##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
## 1. Load packages ----
##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
if(!require(pacman)) install.packages("pacman")
pacman::p_load(tidyverse, janitor, here)

##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
## 2. Import data ----
##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# INSTRUCTION: Here, use `read_csv()` to load in your dataset from the "data" folder in
the "week_02" folder
# Which dataset? The dataset you need should have the same name as your script. Again, i
t is in the "data" subfolder of the "week_02" folder in your pod's RStudio project
# You can find more information about this dataset here:
# https://tinyurl.com/motorcycle-accidents-colombia The dataset is in Spanish, but you c
an use Google and common sense to understand it.









##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
## 3. Create and export a frequency table ----
##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# INSTRUCTION: Using the `tabyl()` function from the {janitor} package,
# make a frequency table of the `REC_MUNRES` variable, which stands for municipal reside
nce
# Then use `write_csv()` to save this table in your "outputs" folder with a descriptive
```

```
name




##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
## 4. Visualize the data to illustrate two key points, then export your plots with code
----
##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# INSTRUCTION: Use {esquisse} to generate two {ggplot2} figures that demonstrate each of the POINTS listed below (If you know how to work with ggplot directly, you can skip esquisse)
# Then use the `ggsave()` function to save your plots in the "outputs" folder with descriptive names

# POINT A: The most common age category (`REC_GRUPO EDAD`) was "20-24" (You will need backticks to refer to this variable)
# POINT B: A majority of female victims were passengers, not drivers (Female passengers have REC_SEXO equal to "Femenino" and passengers have `REC_CONDICION` equal to "Pasajero")

# HINT: The techniques needed above were covered in the "Data dive" and "RStudio Project" lessons.
# With one exception: for POINT B, where you may need to filter the dataset that you are plotting
# Do this by clicking on the Data tab of your esquisse window (bottom right).
# You should see some sliders or variable selectors you can use to filter




##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
## 5. Export the week_02 folder ----
##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# INSTRUCTION: The final step is to export the `week_02` folder from Rstudio cloud.
# Here is an image explaining how to do this: https://imgur.com/a/kbLeIqV
# Then upload the zipped folder as a workshop assignment
# You should only do this AFTER all pod members present in class have finished their own tasks.
# Your final grade will be the average of all the work done in your pod.
# The final deadline is next Tuesday, Oct 25, 23:59 pm CET.



##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
## 6. Present ----
##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
```

```
# INSTRUCTION: Someone from your group will be approached by an instructor and asked to present their work.
# (You will have some time to prepare)
# The selected person will be expected to share their screen, and in about 2 minutes:
# - Say what the given dataset is about. (Show the dataset in your viewer, explain the key variables)
# - Share one of their figures and explain what it is supposed to show
# - (Optional) Explain their answer to the BONUS question below


##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
## BONUS (optional ungraded work): Describing data wrangling steps in words ----
##~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
# For your assigned dataset, try to describe in words the data manipulation steps you would use to achieve the DATA REQUEST below.
# Of course, you don't yet know how to data wrangle in R, so how can you do this?
# You can describe how you would achieve the task:
# - Completely manually (with a printed-out spreadsheet, pen, paper and calculator)
# - With a spreadsheet software like Excel
# - With another data tool that you know (STATA, SPSS)
# - With some combination of the above

# If you have time, try to see if you can actually figure out the answer!

# During the presentation session, a GRAPH instructor will demo how to do this in R
# The goal here is simply to start to get you familiar with some data lingo and concepts

# DATA REQUEST: Find out if the number of accidents in Medellin was increasing or decreasing between 2012 and 2015
```