HW 03 - Car accidents

Figure 1: Photo by Clark Van Der Beken on Unsplash

In this assignment we'll look at traffic accidents in Phoenix. The data are made available on Kaggle online by a scientist at Lyft named Sobhan Moosavi. We will see more on Kaggle throughout the course as it's a great way to find public datasets.

The data covers all recorded car accidents in Phoenix, Arizona from January 2016 to January 2022. Note that some of the variables were modified for the purpose of this assignment.

The data you find on Kaggle includes all accidents across the United States for six years (almost 3 million cases!). You will also find the complete set of variables described on Kaggle. If you haven't set up an account, you really should give it a go.

Getting started

First, open the R Markdown document `hw-03.Rmd` and knit it. Make sure it compiles without errors. The output will be in the file markdown `.md` file with the same name.

Warm up

Before we introduce the data, let's warm up with a simple exercise.

- Update the YAML, changing the author name to your name, and **knit** the document.

Packages

We'll use the **tidyverse** package for much of the data wrangling and visualization. These packages are already installed for you. You can load them by running the following in your Console:

```
library(tidyverse)
```

Data

The data can be found in a separate folder called `data`, and it's called `phoenix_accidents.csv`.

Because our .csv file is in the `data` folder, note that we need to include that in the file path name.

We now need to load in the data before we can do anything with it. We use the function `read_csv()` to do so.

```
phx_accidents <- read_csv("data/phoenix_accidents.csv")
```

The variable descriptions are as follows:

Variable Name	Description
`id`	Accident ID
`latitude`	Latitude of accident location

Variable Name	Description
`longitude`	Longitude of accident location
`severity`	Shows the severity of the accident in terms of traffic delay: Minor/Slight, Moderate, Severe
`road_length`	The length of the road extent affected by the accident (in miles)
`dateof`	Date of the accident
`day_of_week`	Day of the week of the accident
`month_of_year`	Month of the year of the accident
`week_of_year`	Week of the year of the accident
`time`	Time of the accident on the 24h clock
`zipcode`	Phoenix area zipcode
`highway`	Whether the accident occurred on a city street or on a highway
`street`	Shows more detailed information on street/highway location in address field
`side`	Shows the relative side of the street (Right/Left) in address field.
`special_cond`	Special condition that was recorded for the location of the accident: Traffic light/stop sign, Pedestrian crossing, Road/highway junction, Other, Unknown
`light`	Light condition at the time of accident: Daylight, Darkness (artificial lights lit), Darkness (artificial lights unlit), Other
`weather`	Weather condition at the time of accident: Fine, Raining, Snowing, Fog or smoke or dust, Other, Unknown
`wind`	High winds (greater than 26 mph): Windy or Other
`wind_speed`	Shows wind speed (in miles per hour).
`precipitation`	Shows precipitation amount in inches, if there is any.
`visibility`	Shows visibility (in miles).

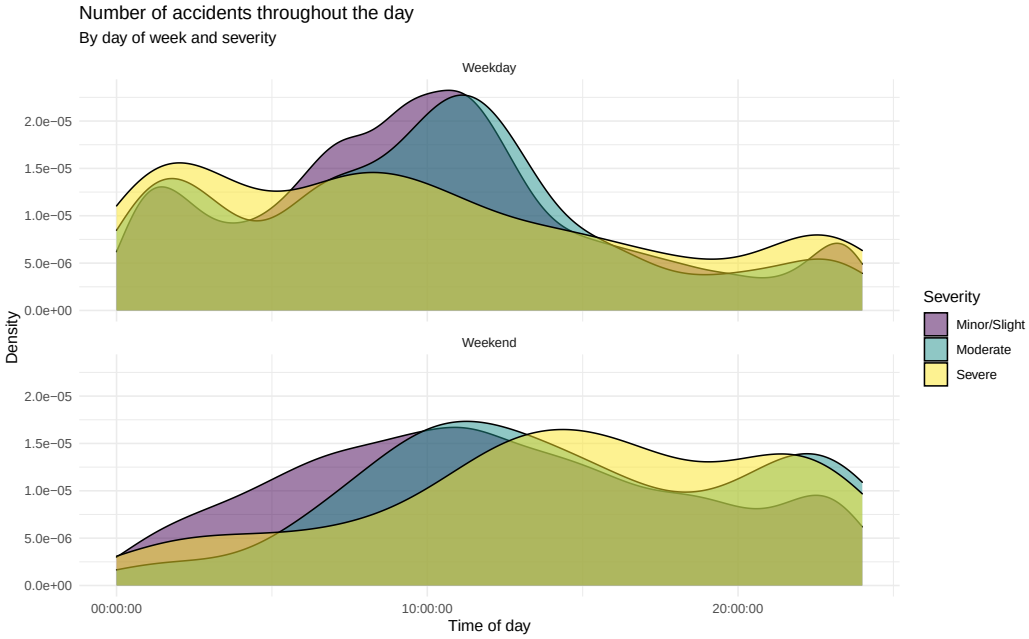
Exercises

1. How many observations (rows) does the dataset have? Instead of hard coding the number in your answer, use inline code.

2. Run `View(az_accidents)` in your Console to view the data in the data viewer. What does each row in the dataset represent?

3. Recreate the following plot, and describe in context of the data. In

your answer, don't forget to label your R chunk as well (where it says `label-me-1`). Your label should be short, informative, shouldn't include spaces, and shouldn't shouldn't repeat a previous label.



4. Create another data visualization based on these data and interpret it. You can choose any variables and any type of visualization you like, but it must have at least three variables, e.g. a scatterplot of x vs. y isn't enough, but if points are colored by z, that's fine. In your answer, don't forget to label your R chunk as well (where it says `label-me-2`).