Lab1

Yonnatan Lourie, Eitan Zimmerman

3/19/2022

- 1. Graph Critique
- 2. Reproducing these analyses
- 3. Freestyle analysis
- 4. Graphical Lineup - check the article
- 5. PCA

Converting dates from string to Date type

```
df$D_month <- do.call(
  recode,
  c(list(df$Month), setNames(c(10,7,12,3,4,11,1,9,6,8,5,2), c("יולי" ,"אוקטובר"
,"פברואר","מאי" ,"אוגוסט" ,"יוני" ,"ספטמבר" ,"ינואר" ,"נובמבר" ,"אפריל" ,"מרס" ,"דצמבר""
) ))
)
```

1. Graph Critique

a. What questions / stories the graphic is trying to answer?

Graph 1 - The graph trying to answer the question if the number of accidents per months in Haifa metropolin is changing? Does the pattern is completely random? or we can infer some insights about particular months or years. We also can see pretty clearly which months our above or below the average.

Graph 2 - The graph tries to show three main things - the relationship between religion, age and road accidents. We can see pretty clearly that in the arab cities there are more accidents by minors (in comparison to the other cities).

b. Do they answer successfully?

I think they answer successfully, but they are not descriptive enough, the second graph need a few seconds to analyze it.

c. Do they raise new questions not addressed?

Yes, for the first graph we can ask which types of accidents occur on this years? what was the weather? How big the population become through the years, etc. For the second graph we will also want to ask which accidents was in each one of the cities. Maybe we can ask in which roads this accidents happen, etc.

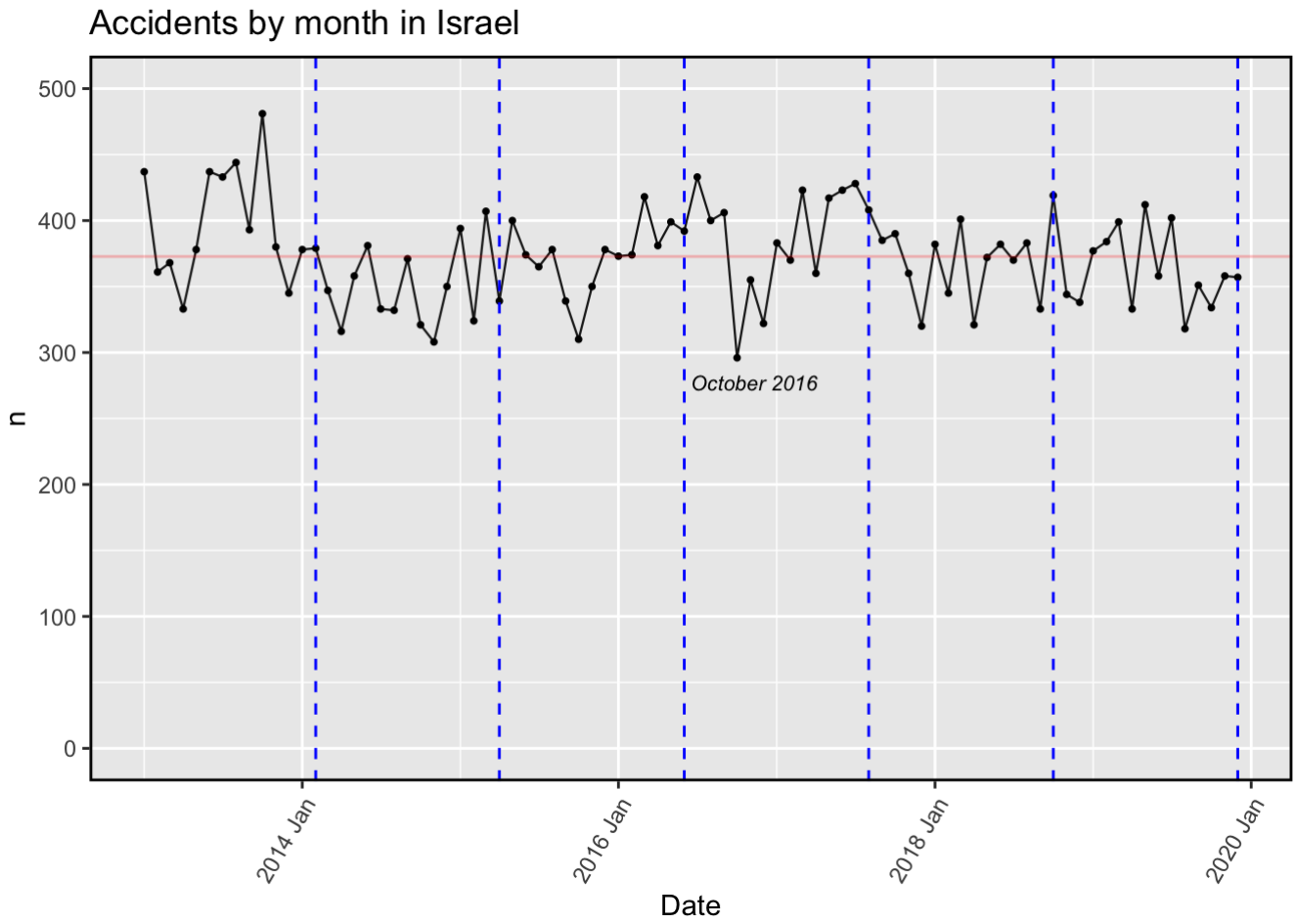
d. Please suggest one way in which these figures can be improved

For the first graph i would plot the number of days with rain for each year, and will split the line for 2 lines - deaths and injured.

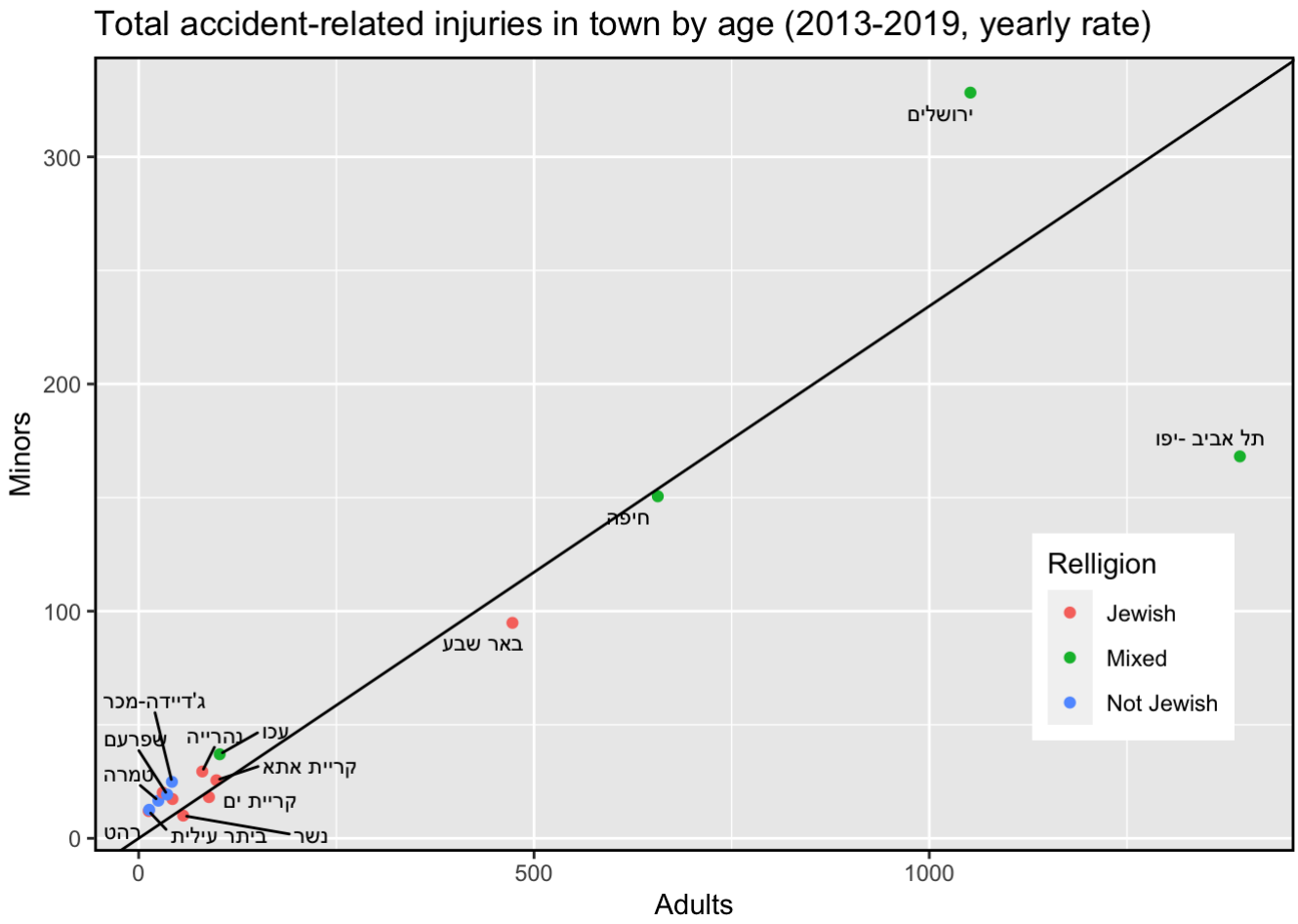
For the second line i would split the graph for more graphs by type to see maybe why we have more accidents with minors at the arab settlements.

2. Reproducing these analyses

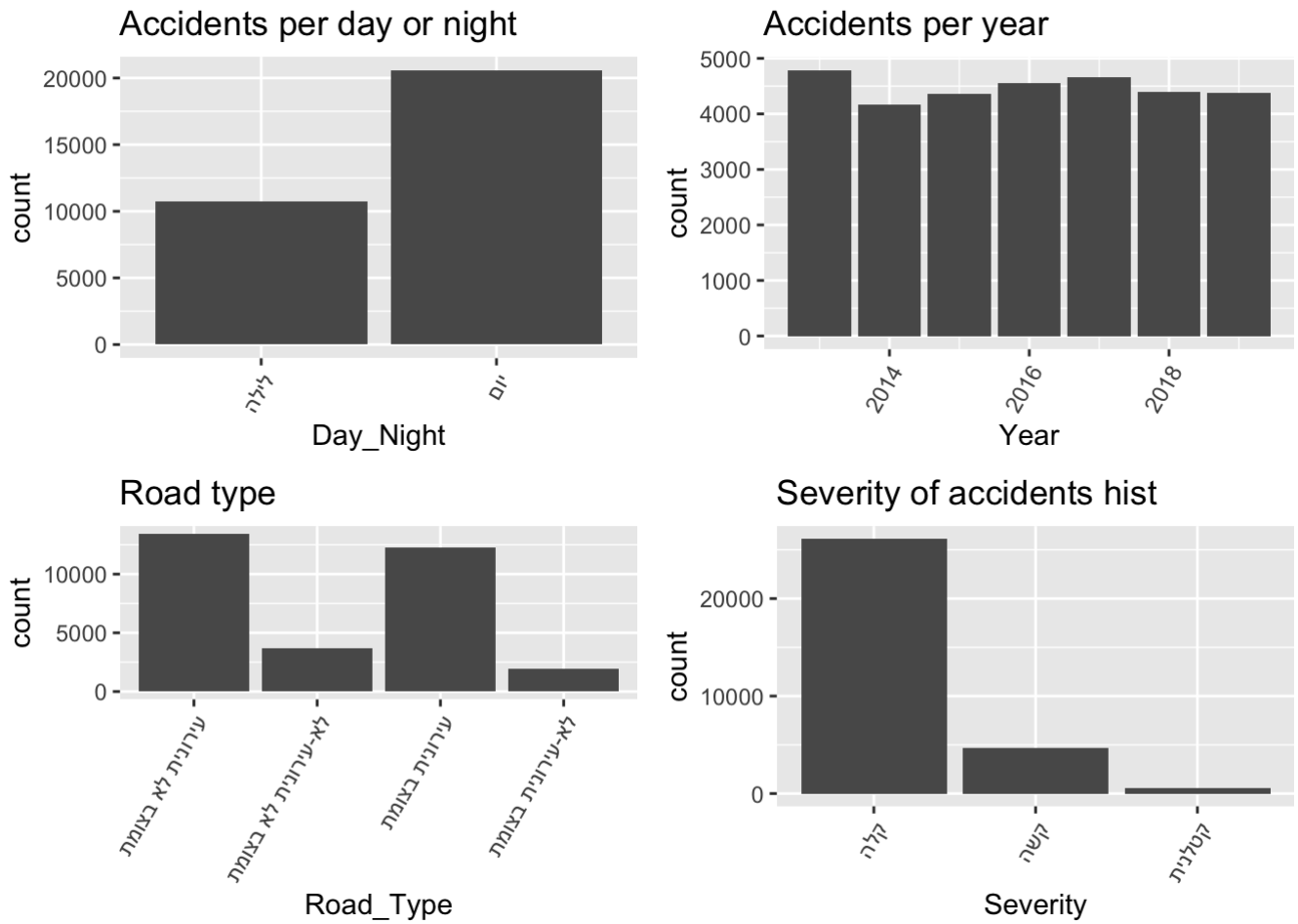
a. A graphic summarizing the total number of accidents by month showing the yearly cycles.



b. A graphic comparing per city the average number of yearly injuries to children vs. adults.

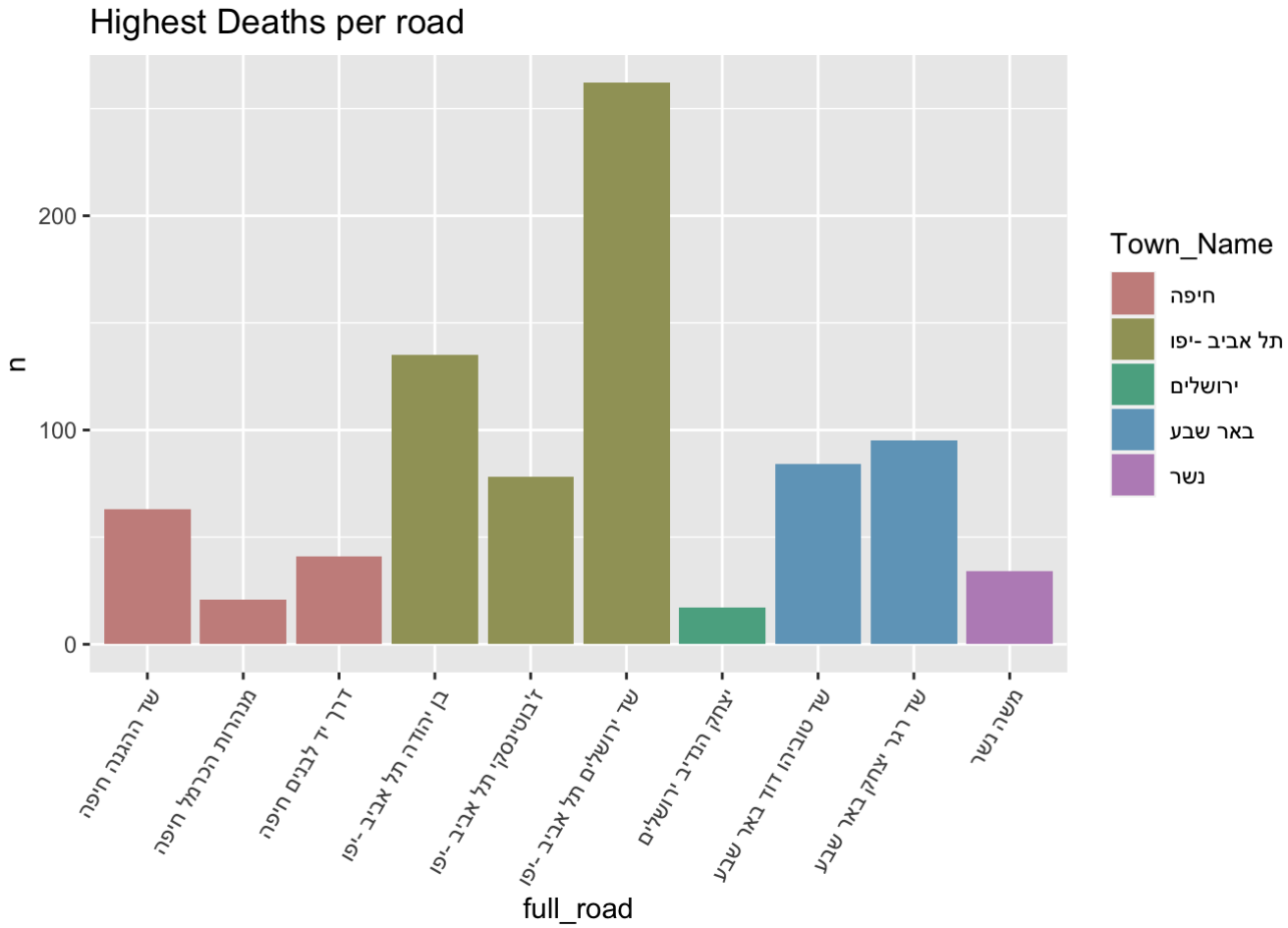


3. Freestyle analysis



From the above graphs we can infer some basic conclusion: - Most of the accidents happen during the day (could be interesting what is the severity ratio of day vs night) . - Most of the accidents are in the city (could be interesting what is the severity ratio of city vs NotCity) - Most of the accidents are with severity light - 2013 was the year with the highest accidents number between 2013-2020.

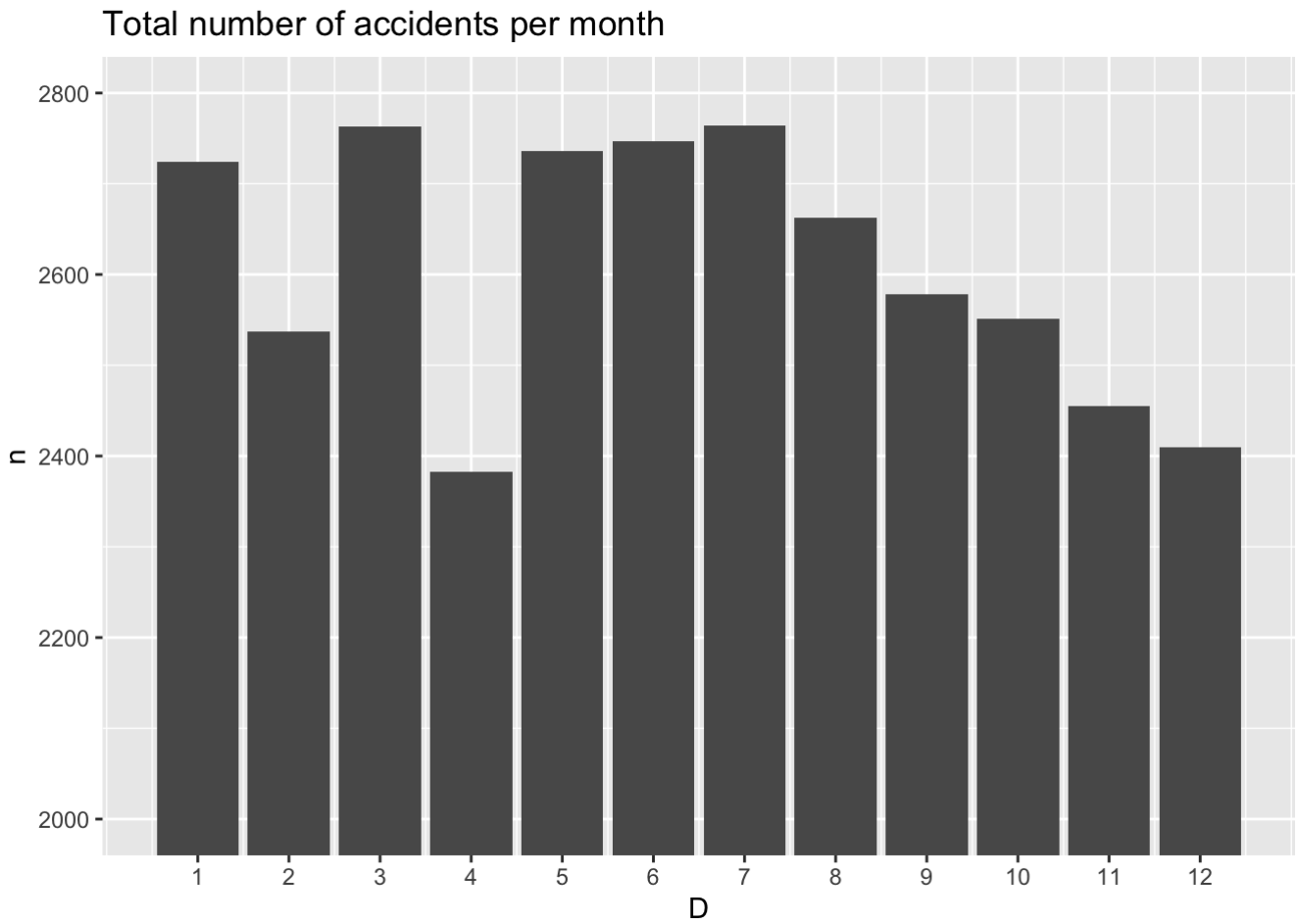
It will be interesting to see if there are certain roads with alot of injuries



Those are the top roads with deaths from a car accident across Israel. But this graph is not telling us the whole picture because all most these ways drive millions of people a year. actually we can see one interesting statistic and it about Moshe Street in Nesher. Though this is one of the main roads of north of Haifa, it still not major as roads in Tel Aviv and Jersualem. This could be a good lead about that particular road. Also we can that Tel Aviv have much higher death statistics than the other cities. Jerusalem is the biggest city in israel and we got only one problematic road (maybe the deaths is more diverse in the roads of the city).

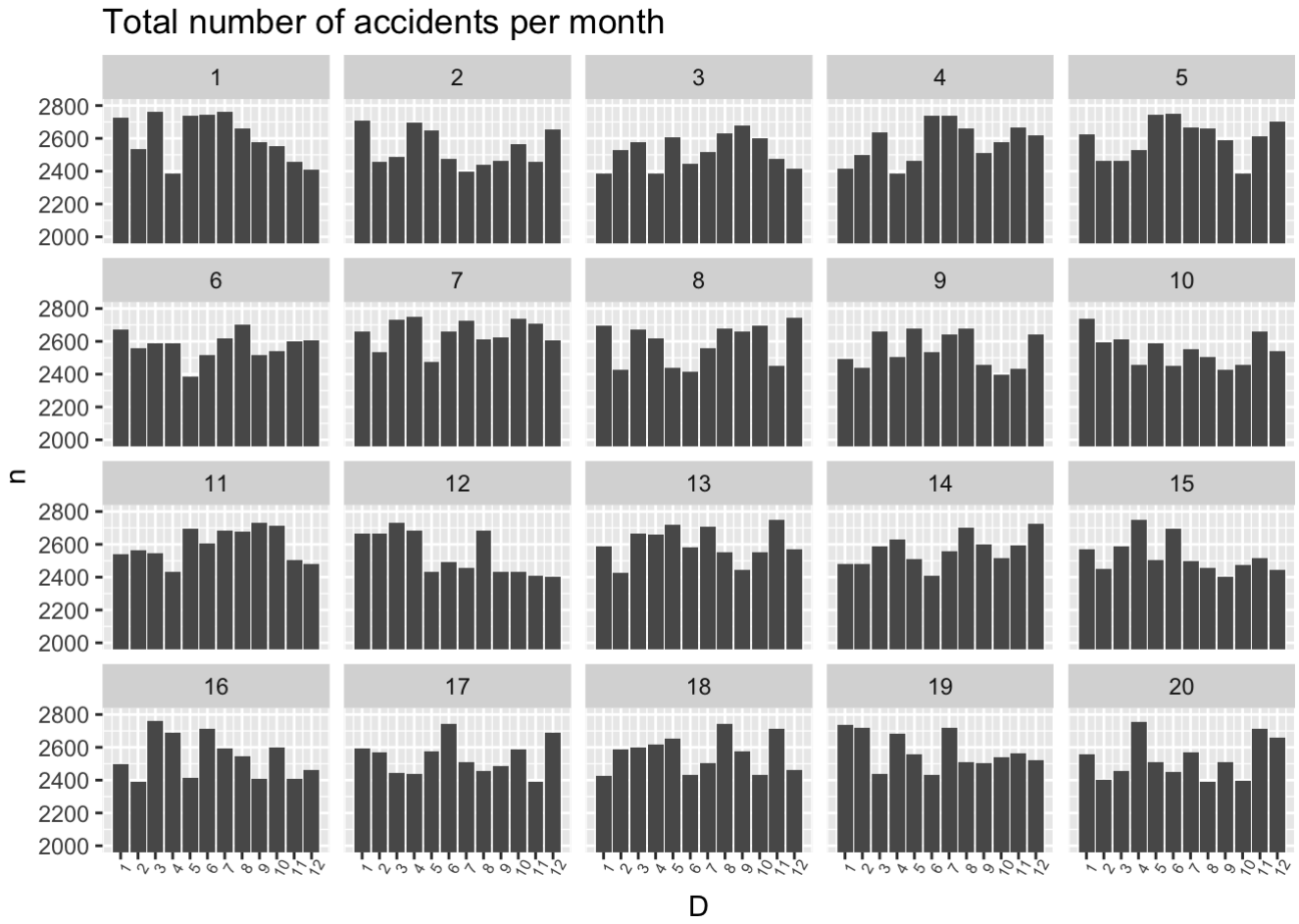
4. Graphical Lineup - check the article

a. Produce a graphic that tries to answer this question from the data. The graph should show the total number of accidents per month summed over the years.



b. Produce 20 simulated data-sets based on the null hypothesis, and produce a graphic for each of them.

We tried to simulate the data with poisson distribution when each month is a poisson random variable, but the graphs was not that similar to the real data. We choose to use a simple uniform distribution.



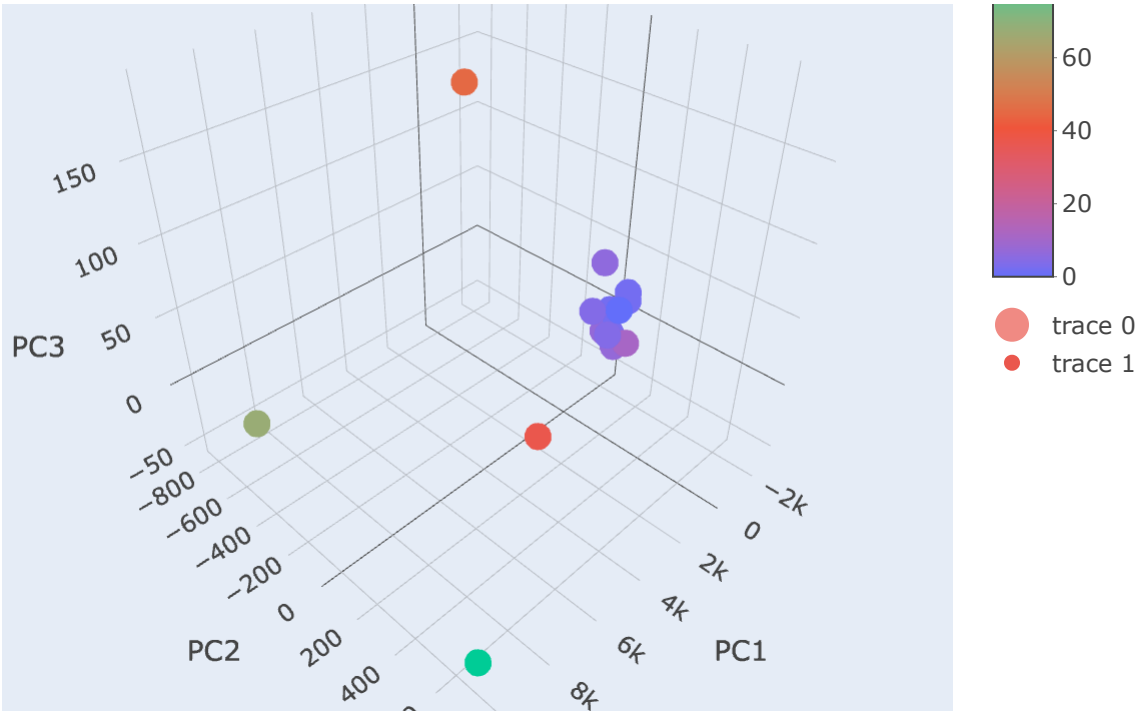
c. Is it easy to tell apart the real data from the simulated ones? How is it different? What have we learned?

We can see that most each month is pretty independent of the months, except from the 1st graph (the real data) which we can see some years the with a monotonic behavior (from May). All the other graphs are pretty independence, and in the real data we can see some lower months (Feb and April). We can reject the null hypothesis because the first graph seems to have a different, more dependent pattern between the months.

5. PCA

```
## No trace type specified:
##   Based on info supplied, a 'scatter3d' trace seems appropriate.
##   Read more about this trace type -> https://plotly.com/r/reference/#scatter
3d
```

```
## No scatter3d mode specifed:
##   Setting the mode to markers
##   Read more about this attribute -> https://plotly.com/r/reference/#scatter-
mode
```



We can see that the smaller towns are clustered together, but the big cities are very far from each other. I dont think that we can say alot about the little towns that are cluster together (because perhpas they are cluster due to their size). But the distance between each one of the big cities is interesting. Probably is due ti the different type of injuries in each one of the cities.