

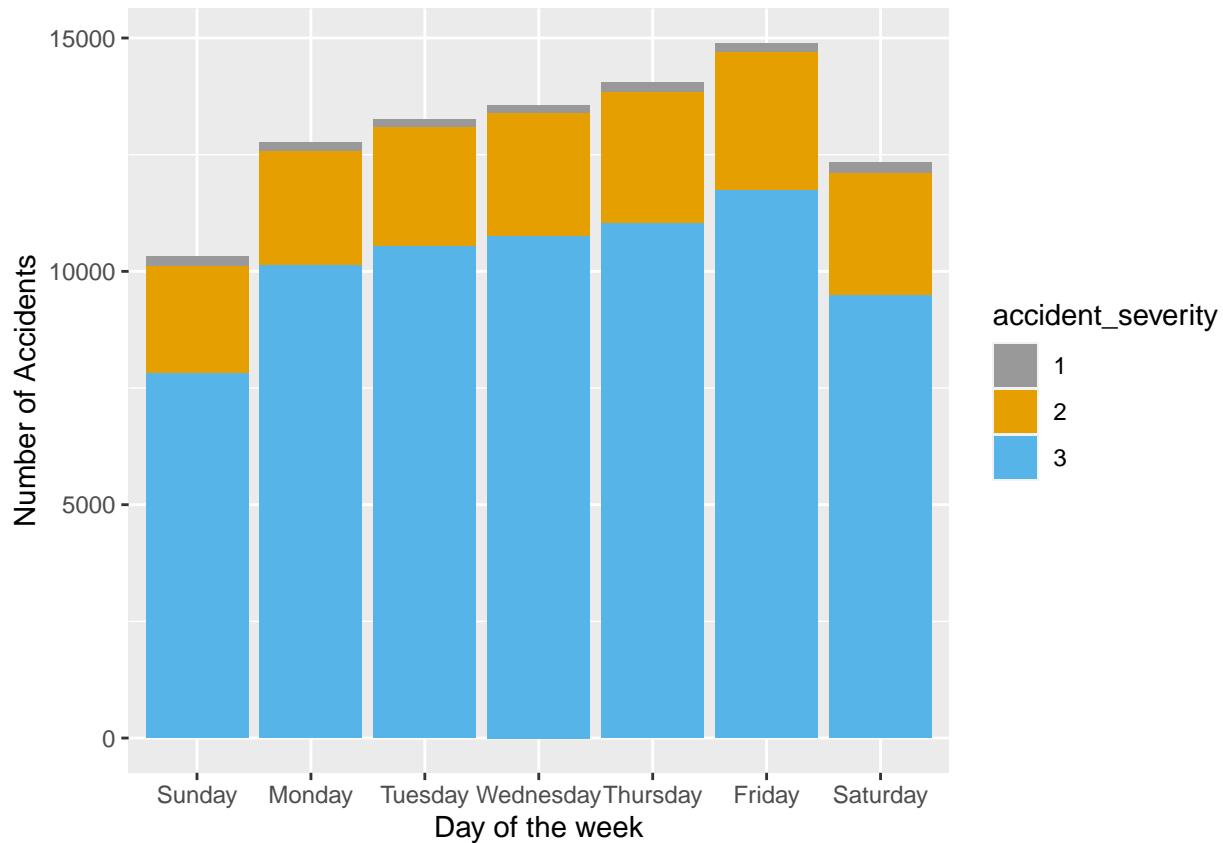
# Preliminary Analysis of RTC Data

Matthew Knowles

2023-03-02

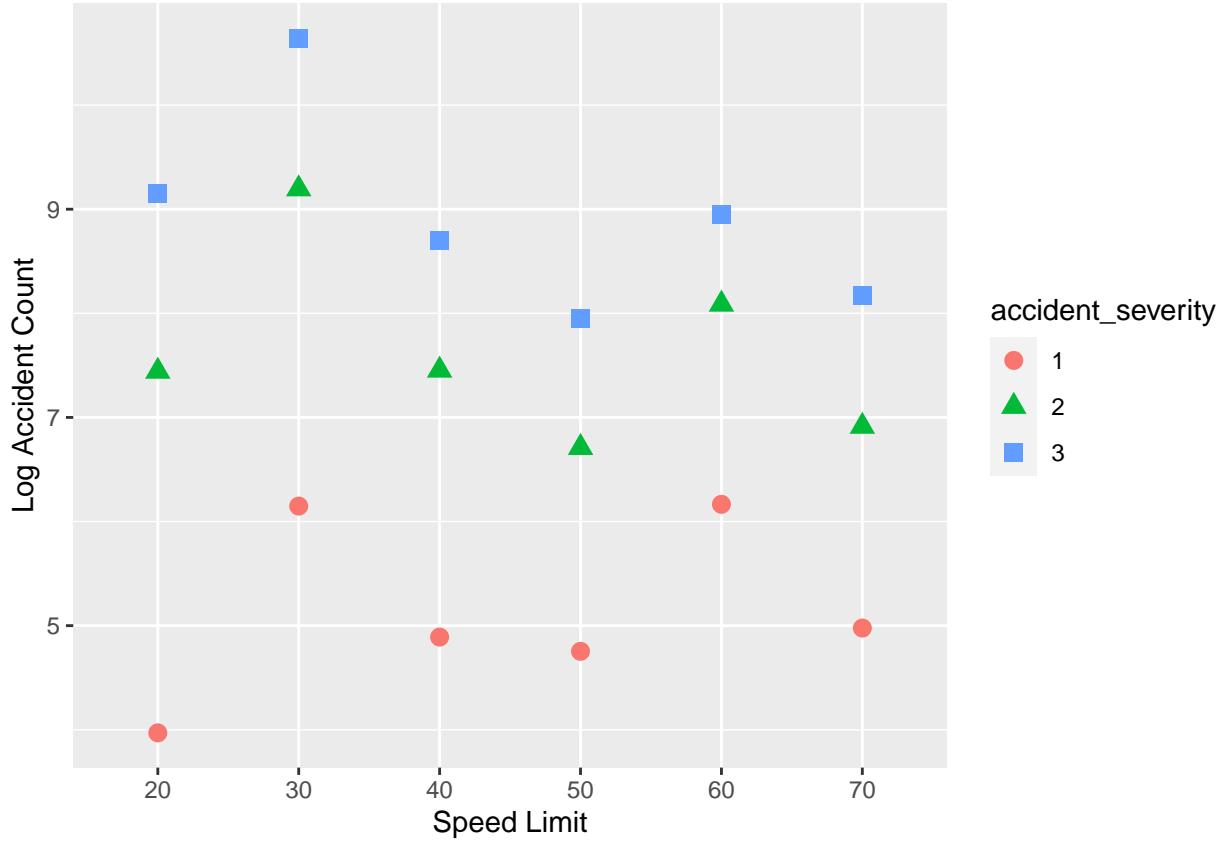
## Right then, lets have a look at this data

On a level, there is so much data here. No one should drive in this country holy shit. I wanted to have a look at whether or not the day of the week would affect anything.



As you can see, during the working week there isn't much difference. It peaks on Friday, I assume when everyone is in a rush to get home from work or get to the pub. Then on saturday and sunday, people are recovering from hangovers so they don't drive as much and there are less accidents.

We then move onto speed limits. I have presented this data in a slightly different way as this will inform what type of link function to use in the GLM.



There doesn't appear to be an immediate pattern in terms of the effect of speed limit on accident count. We will use a standard poisson with log0link model. Not every data is relevant of course, so we will cut down to a few important variables.

Let's take a look at what the model spits out!

```
##
## Call:
## glm(formula = accident_severity ~ ., family = poisson, data = accidents_data)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.42308   0.06348   0.11865   0.14871   0.29355
##
## Coefficients:
##                               Estimate Std. Error z value Pr(>|z|)
## (Intercept)                 1.058e+00  1.438e-02 73.569 < 2e-16 ***
## number_of_vehicles           1.834e-02  2.956e-03  6.205 5.46e-10 ***
## day_of_week                  -1.013e-05 1.033e-03 -0.010 0.992176
## first_road_class             -1.467e-03 1.508e-03 -0.973 0.330715
## road_type                    -4.353e-03 1.244e-03 -3.500 0.000466 ***
## speed_limit                  -1.444e-03 1.616e-04 -8.932 < 2e-16 ***
## junction_detail              4.604e-04 1.624e-04  2.834 0.004593 **
## second_road_class            9.300e-04 7.907e-04  1.176 0.239556
## light_conditions             -3.200e-03 1.171e-03 -2.733 0.006271 **
## weather_conditions           2.449e-03 1.171e-03  2.091 0.036487 *
## road_surface_conditions      3.949e-03 2.338e-03  1.689 0.091241 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##      Null deviance: 7743.6  on 90759  degrees of freedom
## Residual deviance: 7559.3  on 90749  degrees of freedom
## AIC: 270827
##
## Number of Fisher Scoring iterations: 4
```

Some useful plots, perchance?

**density.default(x = model\$residuals)**

