NBA Multilevel Model

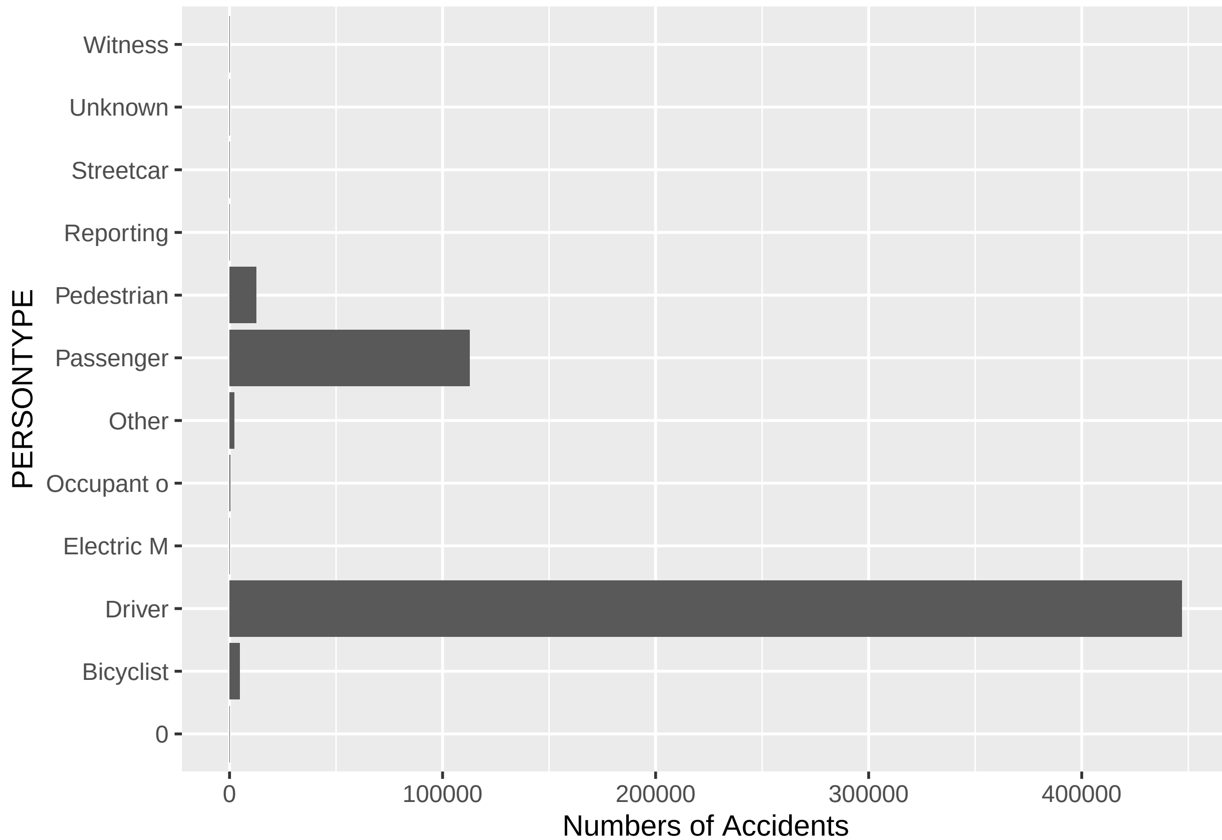
Yuka Chen

2022-12-07

```
df_analysis |>
  group_by(PERSONTYPE) |>
  summarise("Numbers of Accidents"=n()) |>
  knitr::kable()
```

PERSONTYPE	Numbers of Accidents
0	182
Bicyclist	4714
Driver	447058
Electric M	31
Occupant o	338
Other	2200
Passenger	112612
Pedestrian	12438
Reporting	2
Streetcar	4
Unknown	128
Witness	36

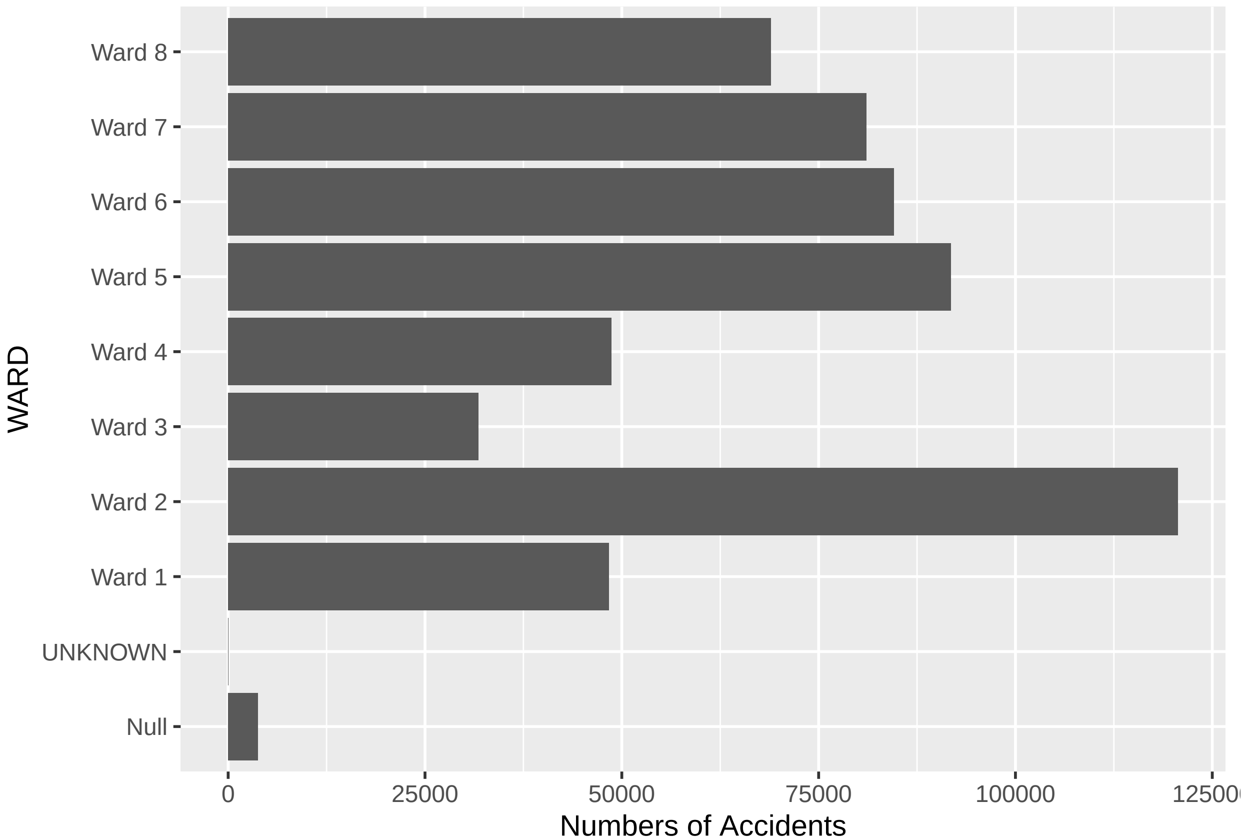
```
options(scipen=10000)

df_analysis |>
  ggplot(aes(x = PERSONTYPE))+
  geom_bar()+
  coord_flip()+
  labs(y = "Numbers of Accidents")+
  scale_y_continuous()
```




```
options(scipen=10000)
df_analysis |>
  group_by(WARD) |>
  summarise("Numbers of Accidents"=n()) |>
  knitr::kable()
```

WARD	Numbers of Accidents
Null	3798
UNKNOWN	2
Ward 1	48351
Ward 2	120658
Ward 3	31802
Ward 4	48680
Ward 5	91812
Ward 6	84594
Ward 7	81096
Ward 8	68950

```
df_analysis |>
  ggplot(aes(x = WARD))+
  geom_bar()+
  coord_flip()+
  labs(y = "Numbers of Accidents")+
  scale_y_continuous()
```




```
#Changing variables to factors
fact_vars <- c('WARD', 'SPEEDING_INVOLVED', 'PERSONTYPE', 'FATAL', 'MAJORINJURY', 'MINORINJURY', 'TICKE
df_analysis[,fact_vars] <- lapply(df_analysis[,fact_vars] , factor)
```

```
#Logistic Regression with the full data set
fatal_logit <- glm(FATAL ~ time_period + month + SPEEDING + IMPAIRED + AGE+ TOTAL_VEHICLES + TOTAL_BICY

logit_sum<-summary(fatal_logit)
logit_sum
```

```
##
## Call:
## glm(formula = FATAL ~ time_period + month + SPEEDING + IMPAIRED +
##     AGE + TOTAL_VEHICLES + TOTAL_BICYCLES + TOTAL_PEDESTRIANS +
##     MAR_SCORE + WARD + TOTAL_GOVERNMENT, family = "binomial",
##     data = df_analysis)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1036  -0.0395  -0.0324  -0.0268   4.3852
##
## Coefficients:
##                        Estimate Std. Error z value             Pr(>|z|)
## (Intercept)          -8.491184   0.939889  -9.034 < 0.0000000000000002 ***
## time_periodEvening   -0.184045   1.000679  -0.184             0.854076
## time_periodMorning    0.139845   1.002458   0.140             0.889053
## time_periodNight      1.348819   0.717811   1.879             0.060235 .
## month2                0.162120   0.274983   0.590             0.555483
```

```
## month3              0.307357   0.255990   1.201             0.229883
## month4              0.265542   0.254681   1.043             0.297112
## month5              0.127202   0.260686   0.488             0.625585
## month6              0.228136   0.257646   0.885             0.375908
## month7              0.436896   0.247667   1.764             0.077725 .
## month8              0.344248   0.252822   1.362             0.173317
## month9              0.044196   0.272607   0.162             0.871207
## month10             0.567618   0.242036   2.345             0.019018 *
## month11             0.077043   0.274933   0.280             0.779306
## month12            -0.152519   0.293617  -0.519             0.603448
## SPEEDINGY           2.719291   0.163182  16.664 < 0.0000000000000002 ***
## IMPAIREDY           0.296374   0.395820   0.749             0.454001
## AGE                 0.010749   0.002215   4.852           0.00000122 ***
## TOTAL_VEHICLES     -0.146731   0.067940  -2.160             0.030795 *
## TOTAL_BICYCLES      0.784156   0.249707   3.140             0.001688 **
## TOTAL_PEDESTRIANS   0.643427   0.043303  14.859 < 0.0000000000000002 ***
## MAR_SCORE          -0.004026   0.001091  -3.690             0.000225 ***
## WARDUNKNOWN        -4.767367 229.127722  -0.021             0.983400
## WARDWard 1         -0.176170   0.617775  -0.285             0.775516
## WARDWard 2         -0.376522   0.600372  -0.627             0.530563
## WARDWard 3         -0.232542   0.636611  -0.365             0.714901
## WARDWard 4          0.352591   0.603358   0.584             0.558964
## WARDWard 5          0.244941   0.595374   0.411             0.680775
## WARDWard 6          0.022380   0.600706   0.037             0.970281
## WARDWard 7          0.607482   0.591938   1.026             0.304769
## WARDWard 8          0.527672   0.596236   0.885             0.376154
## TOTAL_GOVERNMENT   -0.417530   0.194191  -2.150             0.031547 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6782.6  on 579742  degrees of freedom
## Residual deviance: 6404.7  on 579711  degrees of freedom
## AIC: 6468.7
##
## Number of Fisher Scoring iterations: 11
```

```
tidy(fatal_logit) |>
  filter(p.value<0.05) |>
  mutate(p.value = round(p.value, 5)) |>
  dplyr::arrange(p.value) |>
  knitr::kable()
```

term	estimate	std.error	statistic	p.value
(Intercept)	-8.4911837	0.9398892	-9.034239	0.00000
SPEEDINGY	2.7192910	0.1631818	16.664176	0.00000
AGE	0.0107488	0.0022153	4.852089	0.00000
TOTAL_PEDESTRIANS	0.6434274	0.0433031	14.858703	0.00000
MAR_SCORE	-0.0040256	0.0010911	-3.689527	0.00022
TOTAL_BICYCLES	0.7841557	0.2497068	3.140306	0.00169
month10	0.5676182	0.2420357	2.345183	0.01902
TOTAL_VEHICLES	-0.1467311	0.0679403	-2.159707	0.03080

term	estimate	std.error	statistic	p.value
TOTAL_GOVERNMENT	-0.4175302	0.1941906	-2.150105	0.03155

```r
glance(fatal_logit) |>
  knitr::kable()
```

null.deviance	df.null	logLik	AIC	BIC	deviance	df.residual	nobs
6782.645	579742	-3202.342	6468.684	6829.335	6404.684	579711	579743

```r
logout1 <- augment(fatal_logit)
```