260 Project EDA

Group Name: K-Nearest Tailgaters

2021-10-31

```
injuries = read.csv("injuries.csv")
nfl_roster = read.csv("nfl_roster.csv")
```

Check for duplicates, NA's in outcome

```
any(duplicated(injuries[["Full_Name"]])) # no duplicate names (multiple injuries during the year)
```

```
## [1] FALSE
```

```
any(is.na(injuries[["Injury"]])) # no NA's in outcome
```

```
## [1] FALSE
```

`Injury.Status`: What to do with non-football injuries (NFI-R), or COVID? How can we sort based on injury severity?

```
table(injuries$Injury.Status)
```

```
## 
##            Did Not Practice on Thursday. Doubtful for Week 8 at Chicago 
##                                                                        1 
##             Did Not Practice on Thursday. Doubtful for Week 8 vs. Miami 
##                                                                        1 
##      Did Not Practice on Thursday. Doubtful for Week 8 vs. Philadelphia 
##                                                                        1 
##         Did Not Practice on Thursday. Doubtful for Week 8 vs. Tampa Bay 
##                                                                        1 
##        Did Not Practice on Thursday. Doubtful for Week 8 vs. Washington 
##                                                                        1 
##        Did Not Practice on Thursday. Questionable for Week 8 at Atlanta 
##                                                                        1 
##        Did Not Practice on Thursday. Questionable for Week 8 at Chicago 
##                                                                        5 
##         Did Not Practice on Thursday. Questionable for Week 8 at Denver 
##                                                                        2 
##        Did Not Practice on Thursday. Questionable for Week 8 at Detroit 
##                                                                        1 
##        Did Not Practice on Thursday. Questionable for Week 8 at Houston 
##                                                                        4 
##    Did Not Practice on Thursday. Questionable for Week 8 at Kansas City 
##                                                                       4
##      Did Not Practice on Thursday. Questionable for Week 8 at N.Y. Jets
##                                                                       1
##    Did Not Practice on Thursday. Questionable for Week 8 at New Orleans
##                                                                       2
##        Did Not Practice on Thursday. Questionable for Week 8 at Seattle
##                                                                       1
##    Did Not Practice on Thursday. Questionable for Week 8 vs. Cincinnati
##                                                                       1
##        Did Not Practice on Thursday. Questionable for Week 8 vs. Dallas
##                                                                       1
##   Did Not Practice on Thursday. Questionable for Week 8 vs. N.Y. Giants
##                                                                       1
##    Did Not Practice on Thursday. Questionable for Week 8 vs. Pittsburgh
##                                                                       2
##     Did Not Practice on Thursday. Questionable for Week 8 vs. Tampa Bay
##                                                                       3
##     Did Not Practice on Thursday. Questionable for Week 8 vs. Tennessee
##                                                                       1
##    Did Not Practice on Thursday. Questionable for Week 8 vs. Washington
##                                                                       1
##                                                     IR. Injured Reserve
##                                                                     120
##                          IR. Injured Reserve. Expected Return - Week 10
##                                                                      16
##                          IR. Injured Reserve. Expected Return - Week 11
##                                                                      16
##                          IR. Injured Reserve. Expected Return - Week 12
##                                                                       1
##                          IR. Injured Reserve. Expected Return - Week 14
##                                                                       1
##                          IR. Injured Reserve. Expected Return - Week 15
##                                                                       1
##                          IR. Injured Reserve. Expected Return - Week 16
##                                                                       1
##                          IR. Injured Reserve. Expected Return - Week 17
##                                                                       1
##                           IR. Injured Reserve. Expected Return - Week 8
##                                                                      68
##                           IR. Injured Reserve. Expected Return - Week 9
##                                                                      38
##                                                  IR. Reserve - COVID-19
##                                                                       5
##                                       IR. Reserve - Non Football Injury
##                                                                       1
##        Limited Practice on Thursday. Questionable for Week 8 at Atlanta
##                                                                       1
##        Limited Practice on Thursday. Questionable for Week 8 at Buffalo
##                                                                       4
##        Limited Practice on Thursday. Questionable for Week 8 at Chicago
##                                                                       1
##         Limited Practice on Thursday. Questionable for Week 8 at Denver
##                                                                       3
##        Limited Practice on Thursday. Questionable for Week 8 at Detroit
##                                                                       1
##        Limited Practice on Thursday. Questionable for Week 8 at Houston
##                                                                       1
##   Limited Practice on Thursday. Questionable for Week 8 at Indianapolis
##                                                                       2
##    Limited Practice on Thursday. Questionable for Week 8 at Kansas City
##                                                                       2
##  Limited Practice on Thursday. Questionable for Week 8 at L.A. Chargers
##                                                                       5
##      Limited Practice on Thursday. Questionable for Week 8 at Minnesota
##                                                                       2
##    Limited Practice on Thursday. Questionable for Week 8 at New Orleans
##                                                                       2
##        Limited Practice on Thursday. Questionable for Week 8 at Seattle
##                                                                       1
##    Limited Practice on Thursday. Questionable for Week 8 vs. Cincinnati
##                                                                       3
##  Limited Practice on Thursday. Questionable for Week 8 vs. Jacksonville
##                                                                       1
##     Limited Practice on Thursday. Questionable for Week 8 vs. L.A. Rams
##                                                                       2
##   Limited Practice on Thursday. Questionable for Week 8 vs. N.Y. Giants
##                                                                       1
##   Limited Practice on Thursday. Questionable for Week 8 vs. New England
##                                                                       1
##  Limited Practice on Thursday. Questionable for Week 8 vs. Philadelphia
##                                                                       1
##    Limited Practice on Thursday. Questionable for Week 8 vs. Pittsburgh
##                                                                       5
## Limited Practice on Thursday. Questionable for Week 8 vs. San Francisco
##                                                                       2
##     Limited Practice on Thursday. Questionable for Week 8 vs. Tampa Bay
##                                                                       1
##     Limited Practice on Thursday. Questionable for Week 8 vs. Tennessee
##                                                                       1
##    Limited Practice on Thursday. Questionable for Week 8 vs. Washington
##                                                                       1
##      Limited Practice on Wednesday. Questionable for Week 8 vs. Dallas
##                                                                       1
##                                       NFI-R for Week 8 at L.A. Chargers
##                                                                       2
##                                           NFI-R for Week 8 at Minnesota
##                                                                       1
##                                           NFI-R for Week 8 at N.Y. Jets
##                                                                       2
##                                             NFI-R for Week 8 at Seattle
##                                                                       1
##                                         NFI-R for Week 8 vs. Cincinnati
##                                                                       1
##                                             NFI-R for Week 8 vs. Dallas
##                                                                       1
##                                        NFI-R for Week 8 vs. N.Y. Giants
##                                                                       1
##                                       NFI-R for Week 8 vs. Philadelphia
##                                                                       2
##                                          NFI-R for Week 8 vs. Tennessee
##                                                                       1
##                                         NFI-R for Week 8 vs. Washington
##                                                                       1
##                                         NFI-R for Week 9 at N.Y. Giants
##                                                                       1
##                                          NFI-R for Week 9 vs. Minnesota
##                                                                       1
##                Out for Week 8 vs. Cincinnati. Expected Return - Week 10
##                                                                       1
##              Out for Week 8 vs. Jacksonville. Expected Return - Week 11
##                                                                       1
##                 Physically Unable to Perform. Expected Return - Week 11
##                                                                       1
##                 Physically Unable to Perform. Expected Return - Week 12
##                                                                       1
##                  Physically Unable to Perform. Expected Return - Week 8
##                                                                      15
##                                      Questionable for Week 8 at Atlanta
##                                                                       2
##                                    Questionable for Week 8 at Cleveland
##                                                                       2
##                                       Questionable for Week 8 at Denver
##                                                                       2
##                                      Questionable for Week 8 at Detroit
##                                                                       2
##                                      Questionable for Week 8 at Houston
##                                                                       1
##                                 Questionable for Week 8 at Indianapolis
##                                                                       4
##                                  Questionable for Week 8 at Kansas City
##                                                                       2
##                                  Questionable for Week 8 at New Orleans
##                                                                       2
##                                    Questionable for Week 8 vs. Carolina
##                                                                       3
##                                 Questionable for Week 8 vs. N.Y. Giants
##                                                                       1
##                                Questionable for Week 8 vs. Philadelphia
##                                                                       2
##                               Questionable for Week 8 vs. San Francisco
##                                                                       1
##                                   Questionable for Week 8 vs. Tampa Bay
##                                                                       2
##                                   Questionable for Week 8 vs. Tennessee
##                                                                       1
##                                  Questionable for Week 9 at Kansas City
##                                                                       2
##                                  Questionable for Week 9 at N.Y. Giants
##                                                                       4
##                                Questionable for Week 9 at San Francisco
##                                                                       4
##                                   Questionable for Week 9 vs. Minnesota
##                                                                        4
```

One idea: IR > Doubtful > Questionable, then remove NFI and COVID

```r
var_change = function(x) {
  ordinal_injury = c()
  for(i in seq_along(x)) {
    if (str_detect(x[i], "NFI") | str_detect(x[i], "Non Football Injury")) {
      ordinal_injury[i] = "NFI"
    }
    else if (str_detect(x[i], "Questionable")) {
      ordinal_injury[i] = "Questionable"
    }
    else if (str_detect(x[i], "Doubtful")) {
      ordinal_injury[i] = "Doubtful"
    }
    else if (str_detect(x[i], "COVID-19")) {
      ordinal_injury[i] = "COVID-19"
    }
    else if (str_detect(x[i], "IR") | str_detect(x[i], "Physically Unable to Perform")) {
      ordinal_injury[i] = "IR"
    }
    else {
      ordinal_injury[i] = "other"
    }
  }
  ordinal_injury
}

injuries %<>% mutate(ordinal_injury = var_change(injuries$Injury.Status))
table(injuries$ordinal_injury)
```

```
##
##     COVID-19     Doubtful           IR          NFI        other Questionable
##            5            5          280           16            2          117
```

What are the “other” injuries?

```r
injuries %>% filter(ordinal_injury == "other") %>% select(Injury, Injury.Status) # Not the kneecap!!
```

```
##       Injury                                              Injury.Status
## 1 Knee - PCL   Out for Week 8 vs. Cincinnati. Expected Return - Week 10
## 2    Kneecap Out for Week 8 vs. Jacksonville. Expected Return - Week 11
```

Maybe “Out for Week X” == “IR”?

```r
injuries$ordinal_injury[injuries$ordinal_injury == "other"] = "IR"
```

Well, it seems that players are really only "Questionable" or on "IR".

We can remove "COVID-19" and "NFI" injuries, and merge "Doubtful" with "Questionable":

```
injuries %<>% filter(!ordinal_injury %in% c("COVID-19", "NFI")) %>%
              mutate(binary_injury = ifelse(ordinal_injury == "IR", 1, 0))
```

Merge data.frames by "Full_Name"

```
# Some people who are injured are no longer on the roster => out for season
nfl_roster %<>%
  mutate(Team = fix_nfl_names(Team)) %>%
  select(-c(X, College, Drafted, Height, Number, Birthday,
            Draft.Round, Draft.Pick, Birthday_string, Number))

injured_still_on_team = nfl_roster %>%
            inner_join(injuries[, -1], by = c("Full_Name", "Short_Name", "Team"))

roster_with_injuries = nfl_roster %>%
            left_join(injuries[,-1], by = c("Full_Name", "Short_Name", "Team")) %>%
            mutate(ordinal_injury =
                     case_when(str_detect(ordinal_injury, "IR") ~ 2,
                               str_detect(ordinal_injury, "Questionable") ~ 1,
                               str_detect(ordinal_injury, "Doubtful") ~ 1,
                               is.na(ordinal_injury) ~ 0),
                   binary_injury =
                     ifelse(is.na(binary_injury), 0, 1)
                   )
```

NA's per variable

```
apply(injured_still_on_team, 2 , function(x) sum(is.na(x))) # distribution of NA's
```

```
##             Pos          Rating         Ranking          Weight             Age
##               0               2               0               0               0
##            Exp.            Team   height_inches ranking_numeric       Full_Name
##               0               0               0               2               0
##      Short_Name        Position          Injury   Injury.Status            Date
##               0               0               0               0               0
##  ordinal_injury   binary_injury
##               0               0
```

```
apply(roster_with_injuries, 2 , function(x) sum(is.na(x)))
```

```
##             Pos          Rating         Ranking          Weight             Age
##               0              34               0               0               0
##            Exp.            Team   height_inches ranking_numeric       Full_Name
##               0               0               0              34               0
##      Short_Name        Position          Injury   Injury.Status            Date
##               0            2191            2191            2191            2191
##  ordinal_injury   binary_injury
##               0               0
```

Collinearity for Height/Weight

```
cor.test(roster_with_injuries$height_inches,
         roster_with_injuries$Weight) # expected, so will combine to BMI
```

```
##
##  Pearson's product-moment correlation
##
## data:  roster_with_injuries$height_inches and roster_with_injuries$Weight
## t = 50, df = 2508, p-value <2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.686 0.726
## sample estimates:
##   cor
## 0.707
```

Create factors and new variables of interest

```
injured_still_on_team %<>% mutate(Injury = as.factor(Injury),
                     Pos = as.factor(Pos),
                     Team = as.factor(Team))

roster_with_injuries %<>% mutate(Pos = as.factor(Pos),
                     Team = as.factor(Team))


Offensive_Player = c("QB", "RB", "FB", "TB", "HB", "OL", "G", "LG", "RG",
                     "T", "LT", "RT", "C", "WR", "TE")

Defensive_Player = c("DL", "DE", "LE", "RE", "DT", "NT", "LB", "MLB", "ILB",
                     "OLB", "LOLB", "ROLB", "DB", "CB", "S", "SS", "FS")

Special_Teams = c("P", "K", "PR")


injured_still_on_team %<>%
  mutate(Offense = ifelse(Pos %in% c(Offensive_Player, Special_Teams), 1, 0), # offense yes/no
           BMI = (Weight / height_inches^2) * 703) # BMI

roster_with_injuries %<>%
  mutate(Offense = ifelse(Pos %in% c(Offensive_Player, Special_Teams), 1, 0), # offense yes/no
           BMI = (Weight / height_inches^2) * 703) # BMI
```

Descriptive stats

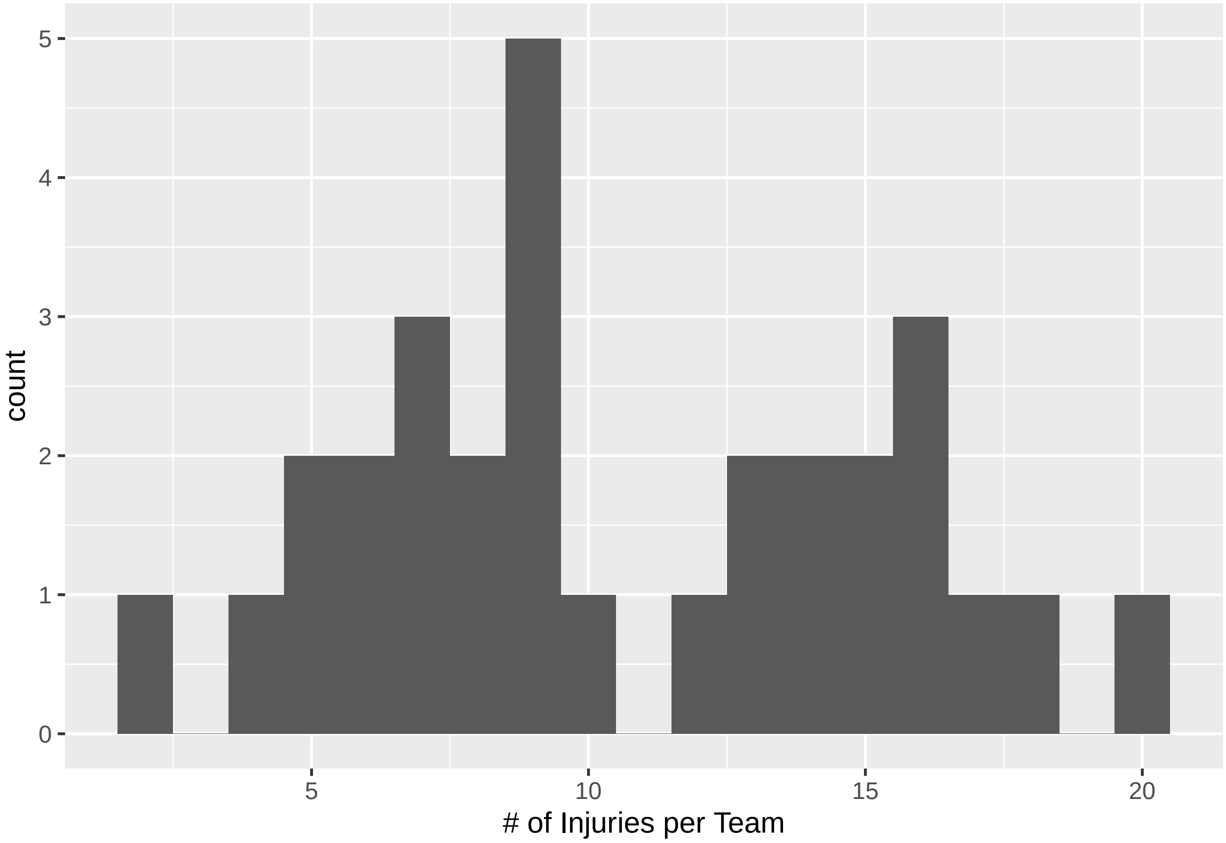
```
number_injury = injured_still_on_team %>% group_by(Team) %>%
  summarize(num_injury = length(Injury))
```

```
pander(summary(number_injury$num_injury)) # ~10 injuries per team
```

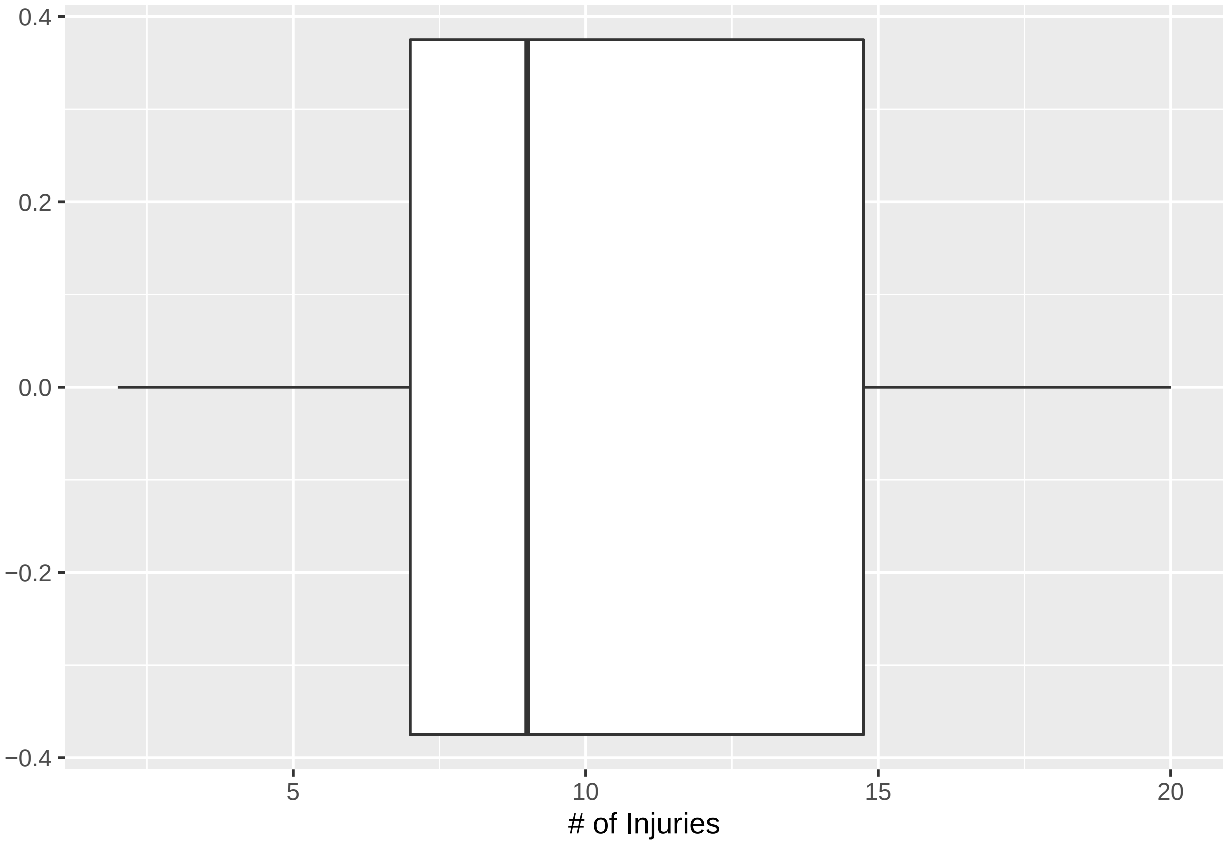
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2	7	9	10.63	14.75	20

Histogram + Boxplot of Injuries per Team

```
number_injury %>%  # histogram (looks kinda bimodal...why??)
  ggplot(aes(num_injury)) +
    geom_histogram(binwidth = 1) +
    xlab("# of Injuries per Team")
```

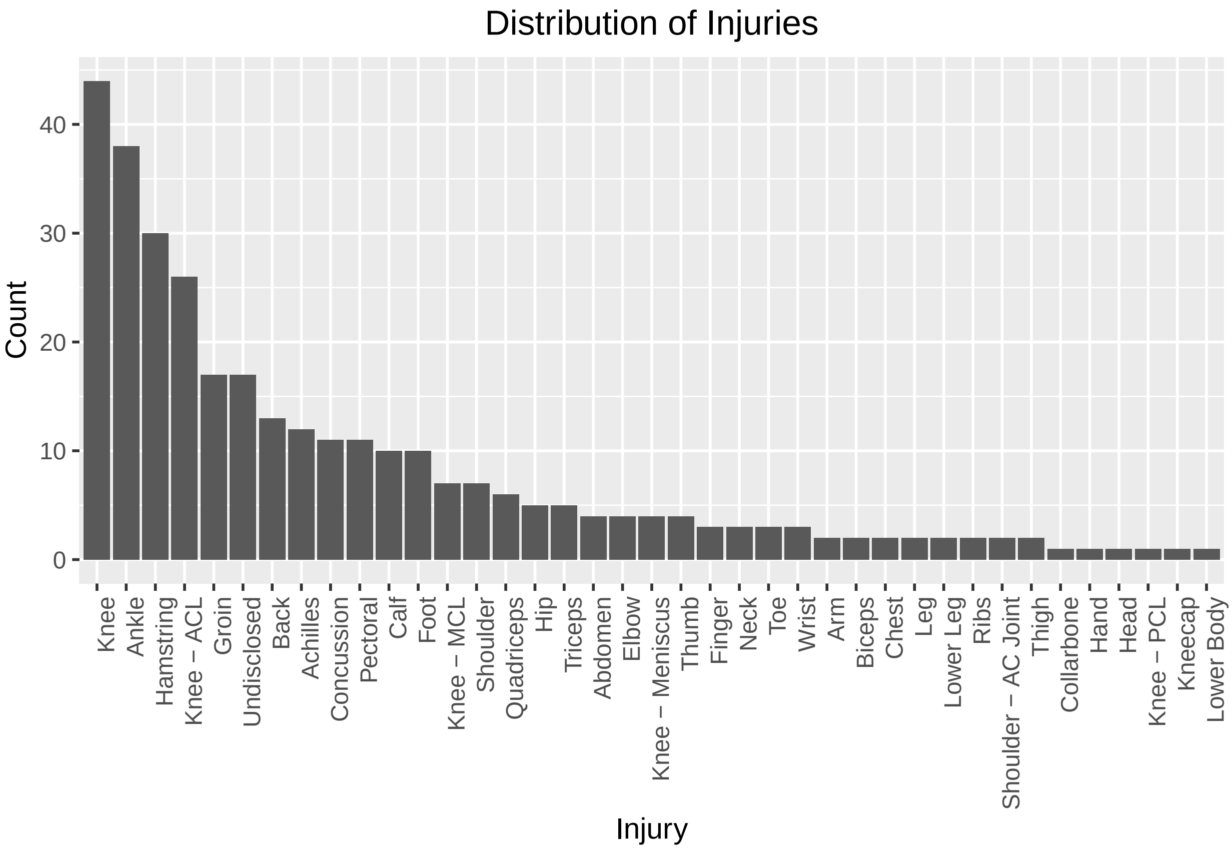


```
number_injury %>% # boxplot (looks kinda symmetric)
  ggplot(aes(num_injury)) +
    geom_boxplot() +
    xlab("# of Injuries")
```

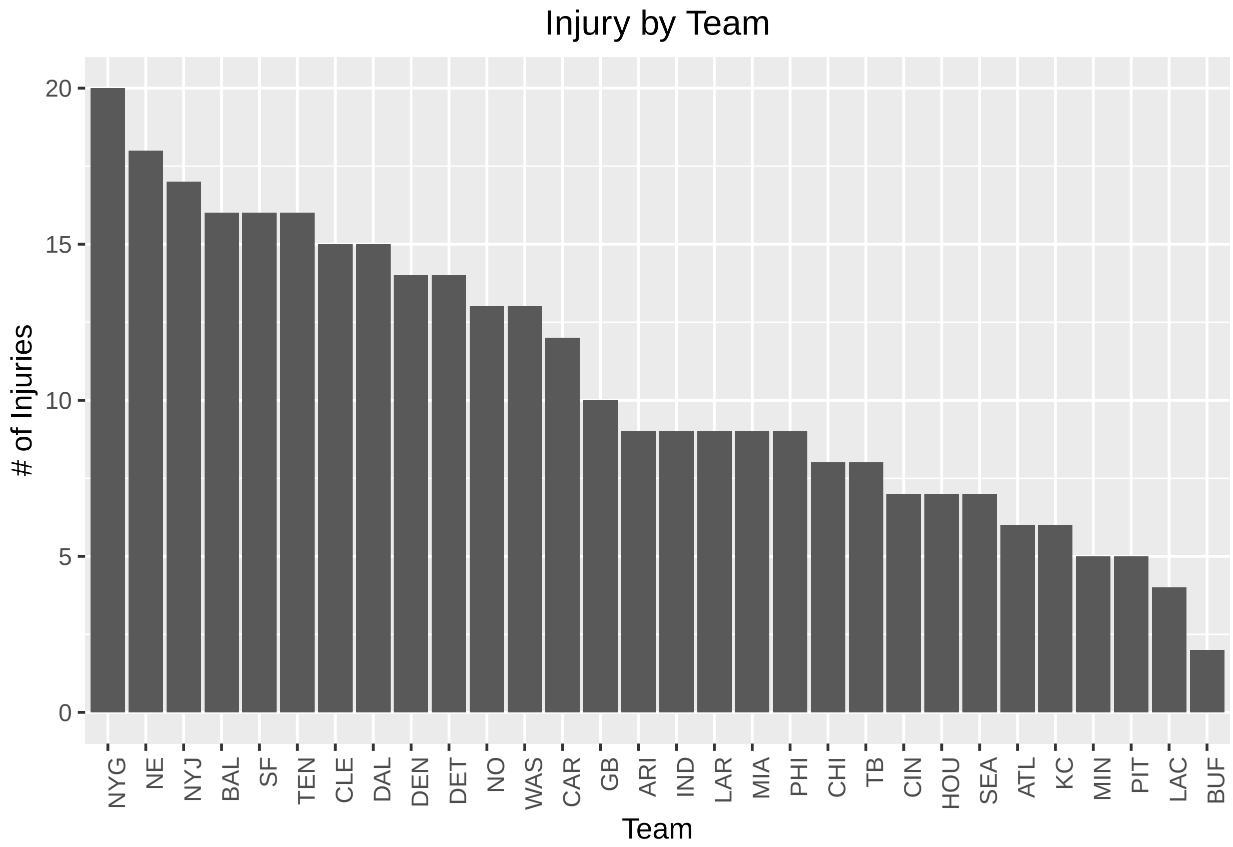


Injuries by Team and Position

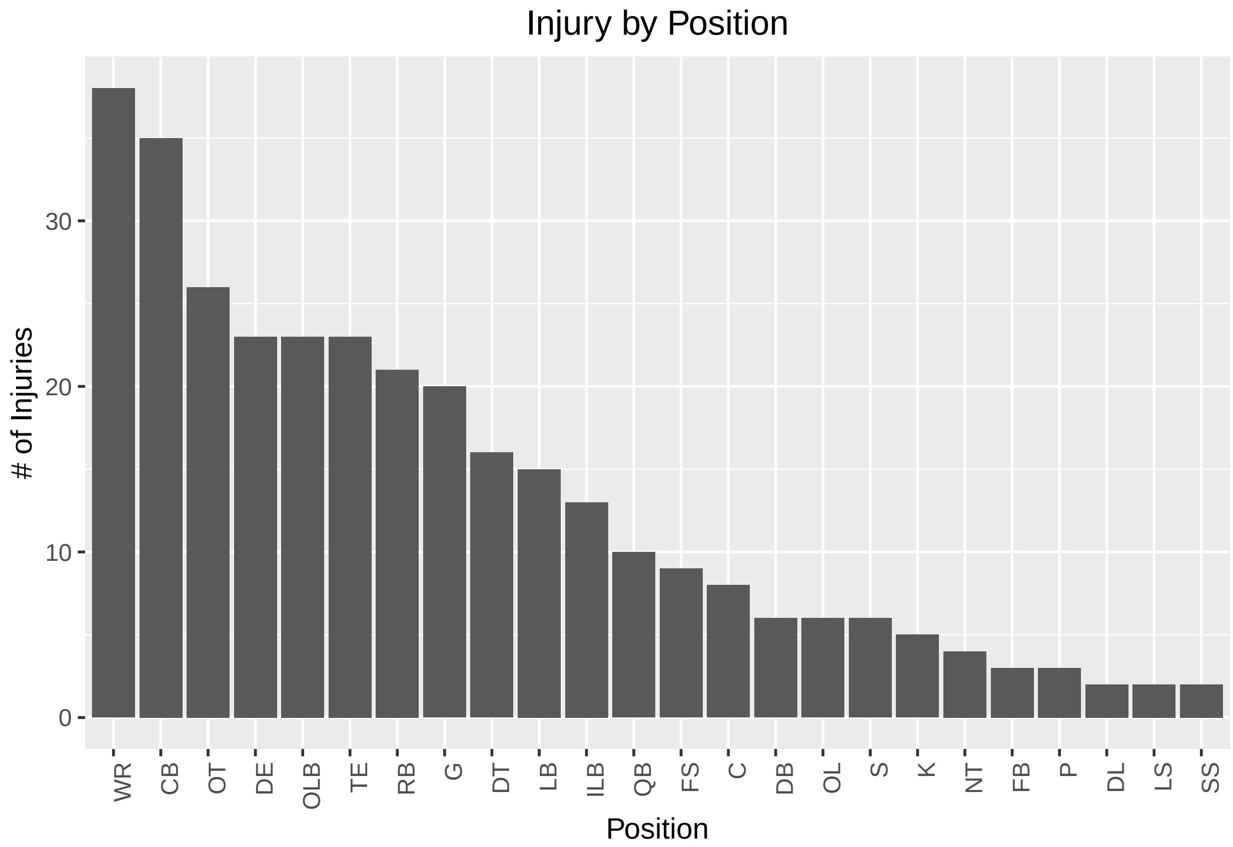
```
injured_still_on_team %>% group_by(Injury) %>%
  summarize(num_injury = length(Injury)) %>%
  mutate(Injury = fct_reorder(Injury, num_injury, .desc = T)) %>%
  ggplot(aes(Injury, num_injury)) +
  geom_col() +
  ylab("Count") +
  ggtitle("Distribution of Injuries") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



```r
# Injuries by Team
injured_still_on_team %>% group_by(Team) %>%
  summarize(num_injury = length(Injury)) %>%
  arrange(desc(num_injury)) %>%
  mutate(Team = fct_reorder(Team, num_injury, .desc = T)) %>%
  ggplot(aes(Team, num_injury)) + geom_col() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  ylab("# of Injuries") +
  ggtitle("Injury by Team")
```



```
# Injuries by Position
injured_still_on_team %>% group_by(Pos) %>%
  summarize(num_injury = length(Injury)) %>%
  arrange(desc(num_injury)) %>%
  mutate(Pos = fct_reorder(Pos, num_injury, .desc = T)) %>%
  ggplot(aes(Pos, num_injury)) +
  geom_col() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
  ylab("# of Injuries") +
  xlab("Position") +
  ggtitle("Injury by Position")
```



Logistic Regression

```
logi_fit = glm(data = roster_with_injuries,
               binary_injury ~ Age + Exp. + BMI + Offense, family = "binomial")

pander(summary(logi_fit))
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.4078	1.128	0.3614	0.7178
Age	-0.1257	0.04862	-2.585	0.009727
Exp.	0.1592	0.0473	3.365	0.0007657
BMI	0.01008	0.01306	0.7723	0.44
Offense	-0.05941	0.1226	-0.4845	0.628

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	1912 on 2509 degrees of freedom
Residual deviance:	1897 on 2505 degrees of freedom

Logistic Regression w/Team + Position

```
logi_fit_Team_Pos = glm(data = roster_with_injuries,
                        binary_injury ~ Team + Pos, family = "binomial")

pander(summary(logi_fit_Team_Pos))
```

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.121	0.5261	-4.032	5.531e-05
TeamATL	-0.447	0.5568	-0.8027	0.4221
TeamBAL	0.5193	0.4548	1.142	0.2536
TeamBUF	-1.504	0.8022	-1.874	0.06089
TeamCAR	0.3422	0.4777	0.7164	0.4737
TeamCHI	-0.1208	0.5193	-0.2327	0.816
TeamCIN	-0.2512	0.5351	-0.4694	0.6388
TeamCLE	0.5823	0.4597	1.267	0.2052
TeamDAL	0.5315	0.459	1.158	0.2469
TeamDEN	0.3628	0.4632	0.7832	0.4335
TeamDET	0.4496	0.4642	0.9686	0.3328
TeamGB	0.1038	0.4938	0.2101	0.8336
TeamHOU	-0.1839	0.5374	-0.3422	0.7322
TeamIND	-0.07163	0.5036	-0.1422	0.8869
TeamJAX	-15.52	455.1	-0.0341	0.9728
TeamKC	-0.4325	0.5626	-0.7687	0.4421
TeamLAC	-0.9017	0.6284	-1.435	0.1513
TeamLAR	0.02505	0.5085	0.04927	0.9607
TeamMIA	0.04944	0.5062	0.09767	0.9222
TeamMIN	-0.5582	0.5866	-0.9516	0.3413
TeamNE	0.8002	0.4492	1.781	0.07486
TeamNO	0.4447	0.4713	0.9436	0.3454
TeamNYG	0.8556	0.4422	1.935	0.05303
TeamNYJ	0.6607	0.4506	1.466	0.1426
TeamOAK	-15.54	441.3	-0.03523	0.9719
TeamPHI	-0.04432	0.5057	-0.08763	0.9302
TeamPIT	-0.5947	0.586	-1.015	0.3102
TeamSEA	-0.2347	0.5352	-0.4385	0.661
TeamSF	0.6574	0.4552	1.444	0.1487
TeamTB	-0.08938	0.5198	-0.172	0.8635
TeamTEN	0.5517	0.4544	1.214	0.2247
TeamWAS	0.4446	0.4701	0.9457	0.3443
PosCB	0.1028	0.4274	0.2405	0.8099
PosDB	1.141	0.6325	1.804	0.0712
PosDE	0.09126	0.4486	0.2034	0.8388
PosDL	-0.7068	0.8338	-0.8477	0.3966
PosDT	0.1353	0.4736	0.2858	0.775
PosFB	-0.1212	0.7373	-0.1644	0.8694
PosFS	0.2382	0.5343	0.4459	0.6557
PosG	0.2502	0.4585	0.5457	0.5853
PosILB	0.2756	0.4928	0.5593	0.5759
PosK	-0.1487	0.6164	-0.2413	0.8093
PosLB	0.04812	0.4792	0.1004	0.92
PosLS	-0.7754	0.8296	-0.9347	0.3499
PosNT	-0.1299	0.6617	-0.1964	0.8443

	Estimate	Std. Error	z value	Pr(>|z|)
PosOL	-0.02354	0.5882	-0.04002	0.9681
PosOLB	0.5943	0.4533	1.311	0.1898
PosOT	0.2513	0.4423	0.5681	0.5699
PosP	-0.3598	0.7225	-0.4981	0.6184
PosQB	-0.2456	0.512	-0.4797	0.6314
PosRB	0.04986	0.4526	0.1102	0.9123
PosS	-0.168	0.5832	-0.288	0.7733
PosSS	-1.443	0.8189	-1.762	0.07812
PosTE	0.1472	0.4479	0.3287	0.7424
PosWR	0.06401	0.4239	0.151	0.88

(Dispersion parameter for binomial family taken to be 1)

Null deviance:	1912 on 2509 degrees of freedom
Residual deviance:	1788 on 2455 degrees of freedom

Maybe too much information in the outcome is lost by making injury binary, maybe ordinal or multinomial would be preferred.

We also might want to use data from past years as we are only half way through the current season.

Multinomial

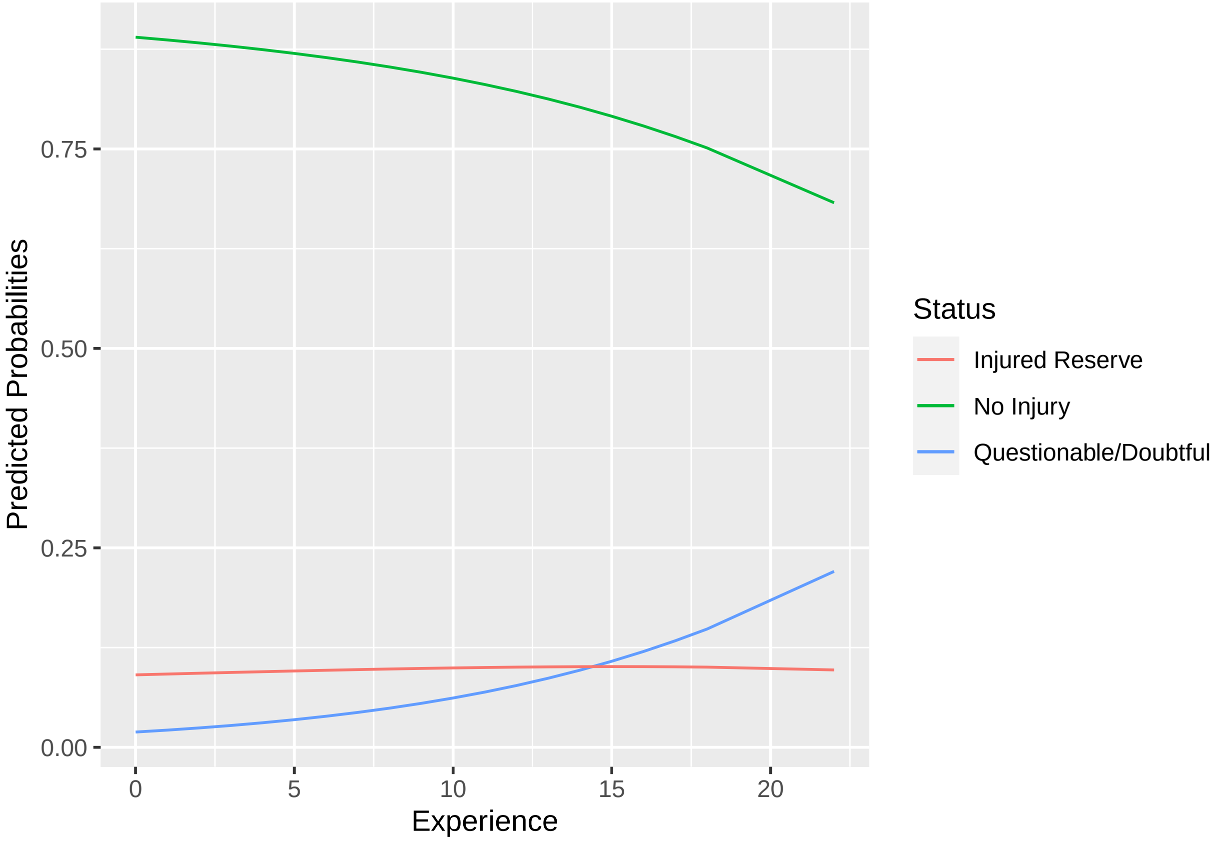
```
multinom_injury = multinom(data = roster_with_injuries,
                           ordinal_injury ~ Exp.) # Experience only
```

```
## # weights:  9 (4 variable)
## initial  value 2757.516845
## iter  10 value 1130.588337
## final  value 1129.843819
## converged
```

```
summary(multinom_injury)
```

```
## Call:
## multinom(formula = ordinal_injury ~ Exp., data = roster_with_injuries)
##
## Coefficients:
##   (Intercept)   Exp.
## 1       -3.84 0.1232
## 2       -2.28 0.0151
##
## Std. Errors:
##   (Intercept)   Exp.
## 1       0.191 0.0297
## 2       0.109 0.0214
##
## Residual Deviance: 2260
## AIC: 2268
```

```
ggplot(roster_with_injuries, aes(Exp., multinom_injury$fitted.values[,1])) +
  geom_line(aes(Exp., multinom_injury$fitted.values[,1], color = "No Injury")) +
  geom_line(aes(Exp., multinom_injury$fitted.values[,2], color = "Questionable/Doubtful")) +
  geom_line(aes(Exp., multinom_injury$fitted.values[,3], color = "Injured Reserve")) +
  scale_color_discrete("Status") +
  ylab("Predicted Probabilities") +
  xlab("Experience")
```


Predicted Probabilities
Experience
Status
Injured Reserve
No Injury
Questionable/Doubtful