

# COPD Analysis

```
#Install/Load Packages & Files
```

```
packages <- c('markdown',"epitools", "ggplot2", "readxl", "dplyr", "tidyr", "knitr", "lubridate",  
  , "MASS", "summarytools", "magrittr", "tidyverse")  
#install.packages(packages)
```

```
#Load Libraries
```

```
lapply(packages, require, character.only = TRUE)
```

```
## Loading required package: markdown
```

```
## Loading required package: epitools
```

```
## Loading required package: ggplot2
```

```
## Loading required package: readxl
```

```
## Loading required package: dplyr
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
## Loading required package: tidyr
```

```
## Loading required package: knitr
```

```
## Loading required package: lubridate
```

```
##
```

```
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':  
##  
##   date, intersect, setdiff, union
```

```
## Loading required package: MASS
```

```
##  
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':  
##  
##   select
```

```
## Loading required package: summarytools
```

```
## Loading required package: magrittr
```

```
##  
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:tidyr':  
##  
##   extract
```

```
## Loading required package: tidyverse
```

```
## — Attaching packages ————— tidyverse 1.3.2 —  
## ✓ tibble 3.1.8      ✓ stringr 1.4.1  
## ✓ readr 2.1.2      ✓ forcats 0.5.2  
## ✓ purrr 0.3.4  
## — Conflicts ————— tidyverse_conflicts() —  
## X lubridate::as.difftime() masks base::as.difftime()  
## X lubridate::date()        masks base::date()  
## X magrittr::extract()      masks tidyr::extract()  
## X dplyr::filter()          masks stats::filter()  
## X lubridate::intersect()   masks base::intersect()  
## X dplyr::lag()             masks stats::lag()  
## X MASS::select()          masks dplyr::select()  
## X purrr::set_names()       masks magrittr::set_names()  
## X lubridate::setdiff()     masks base::setdiff()  
## X lubridate::union()       masks base::union()  
## X tibble::view()          masks summarytools::view()
```

```
## [[1]]
## [1] TRUE
##
## [[2]]
## [1] TRUE
##
## [[3]]
## [1] TRUE
##
## [[4]]
## [1] TRUE
##
## [[5]]
## [1] TRUE
##
## [[6]]
## [1] TRUE
##
## [[7]]
## [1] TRUE
##
## [[8]]
## [1] TRUE
##
## [[9]]
## [1] TRUE
##
## [[10]]
## [1] TRUE
##
## [[11]]
## [1] TRUE
##
## [[12]]
## [1] TRUE
```

```
#Load Data
```

```
COPD <- read_excel("C:/Users/Kelsey/Downloads/COPD.xlsx")
```

```
## New names:
```

```
## • `` -> `...1`
```

```
descr(COPD)
```

```
## Non-numerical variable(s) ignored: COPDSEVERITY, MWT1, MWT2, MWT1Best
```

## ## Descriptive Statistics

## COPD

## N: 101

##

## ...1 AGE AGEquartiles AtrialFib CAT copd Diabetes

FEV1

## -----

-----

## Mean 51.00 70.10 2.48 0.20 19.34 2.20 0.21

1.60

## Std.Dev 29.30 7.90 1.11 0.40 18.67 0.88 0.41

0.67

## Min 1.00 44.00 1.00 0.00 3.00 1.00 0.00

0.45

## Q1 26.00 65.00 1.00 0.00 12.00 2.00 0.00

1.10

## Median 51.00 71.00 3.00 0.00 18.00 2.00 0.00

1.60

## Q3 76.00 75.00 3.00 0.00 24.00 3.00 0.00

1.96

## Max 101.00 88.00 4.00 1.00 188.00 4.00 1.00

3.18

## MAD 37.06 5.93 1.48 0.00 8.90 1.48 0.00

0.68

## IQR 50.00 10.00 2.00 0.00 12.00 1.00 0.00

0.86

## CV 0.57 0.11 0.45 2.02 0.97 0.40 1.96

0.42

## Skewness 0.00 -0.70 0.00 1.49 7.24 0.30 1.42

0.45

## SE.Skewness 0.24 0.24 0.24 0.24 0.24 0.24 0.24

0.24

## Kurtosis -1.24 0.65 -1.36 0.23 62.95 -0.68 0.01

-0.54

## N.Valid 101.00 101.00 101.00 101.00 101.00 101.00 101.00

101.00

## Pct.Valid 100.00 100.00 100.00 100.00 100.00 100.00 100.00

100.00

##

## Table: Table continues below

##

##

##

## FEV1PRED FVC FVCPRED gender HAD hypertension ID

IHD

## -----

-----

## Mean 58.53 2.95 86.44 0.64 11.18 0.12 91.41

0.09

## Std.Dev 22.29 0.98 21.74 0.48 8.59 0.33 51.52

0.29

## Min 3.29 1.14 27.00 0.00 0.00 0.00 1.00

```

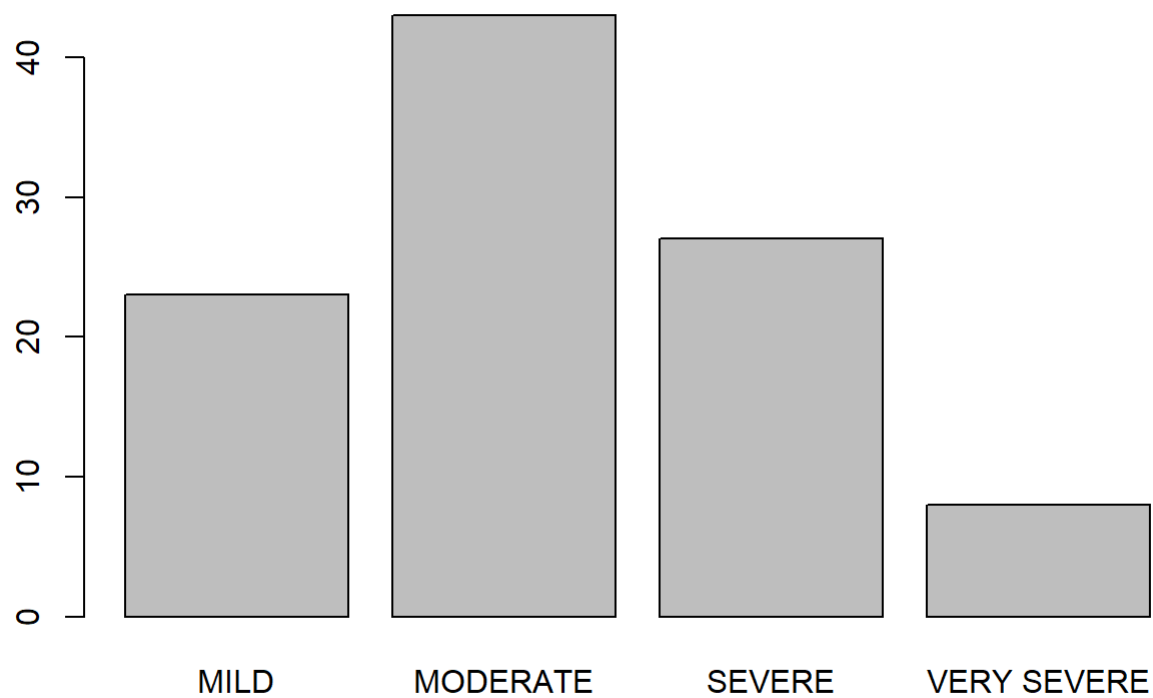
0.00
##          Q1      42.00    2.27    71.00    0.00    6.00          0.00    49.00
0.00
##          Median    60.00    2.77    84.00    1.00   10.00          0.00    87.00
0.00
##          Q3      75.00    3.63   103.00    1.00   15.00          0.00   143.00
0.00
##          Max     102.00    5.37   132.00    1.00   56.20          1.00   169.00
1.00
##          MAD      23.72    1.01    22.24    0.00    7.41          0.00    68.20
0.00
##          IQR      33.00    1.36    32.00    1.00    9.00          0.00    94.00
0.00
##          CV        0.38    0.33     0.25    0.75    0.77          2.74     0.56
3.21
##          Skewness  -0.16    0.49     0.00   -0.59    1.72          2.32   -0.05
2.84
##          SE.Skewness  0.24    0.24     0.24    0.24    0.24          0.24     0.24
0.24
##          Kurtosis  -0.62   -0.58   -0.63   -1.67    5.87          3.42   -1.30
6.14
##          N.Valid   101.00   101.00   101.00   101.00   101.00        101.00   101.00    1
01.00
##          Pct.Valid  100.00   100.00   100.00   100.00   100.00        100.00   100.00    1
00.00
##
## Table: Table continues below
##
##
##
##          muscular    PackHistory    SGRQ    smoking
## -----
##          Mean        0.19          39.70    40.19    1.84
##          Std.Dev      0.39          24.56    18.24    0.37
##          Min          0.00           1.00     2.00    1.00
##          Q1           0.00          20.00    28.41    2.00
##          Median       0.00          36.00    38.21    2.00
##          Q3           0.00          54.00    55.23    2.00
##          Max          1.00         109.00    77.44    2.00
##          MAD          0.00          23.72    18.27    0.00
##          IQR          0.00          34.00    26.82    0.00
##          CV           2.09           0.62     0.45    0.20
##          Skewness     1.57           0.73     0.18   -1.84
##          SE.Skewness   0.24           0.24     0.24    0.24
##          Kurtosis      0.48           0.22    -0.72    1.41
##          N.Valid      101.00         101.00   101.00   101.00
##          Pct.Valid     100.00         100.00   100.00   100.00

```

```

#Barplot of severity
barplot(table(COPD$COPDSEVERITY))

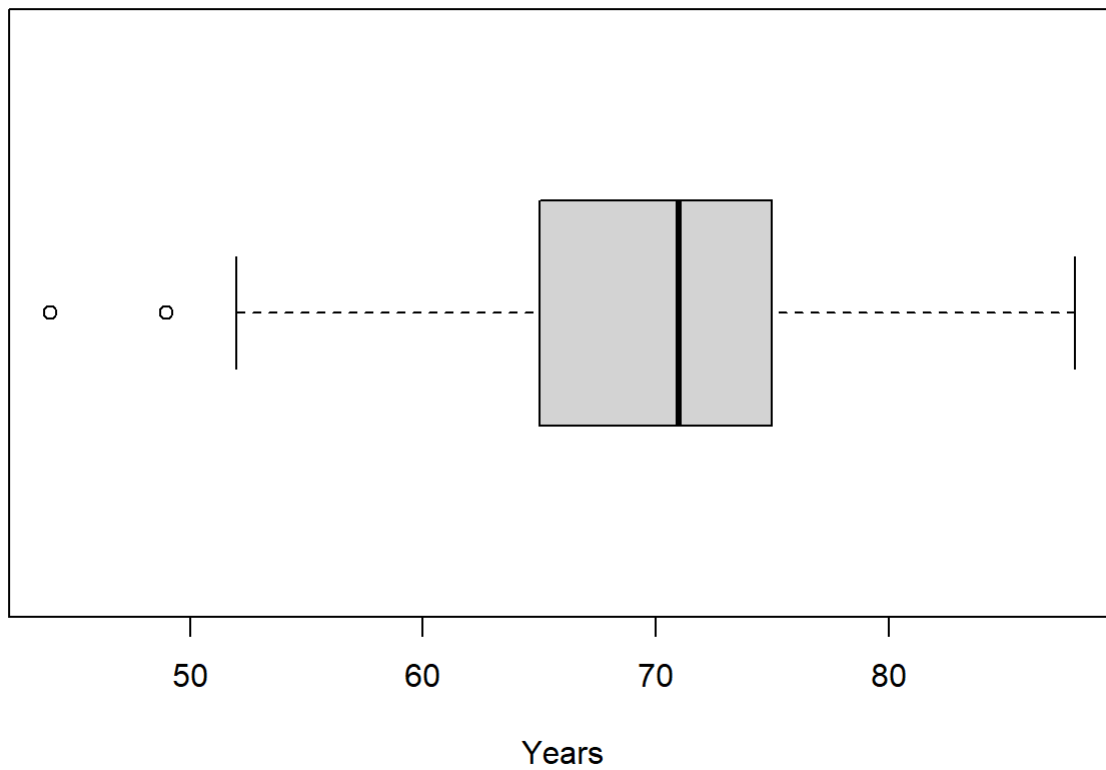
```



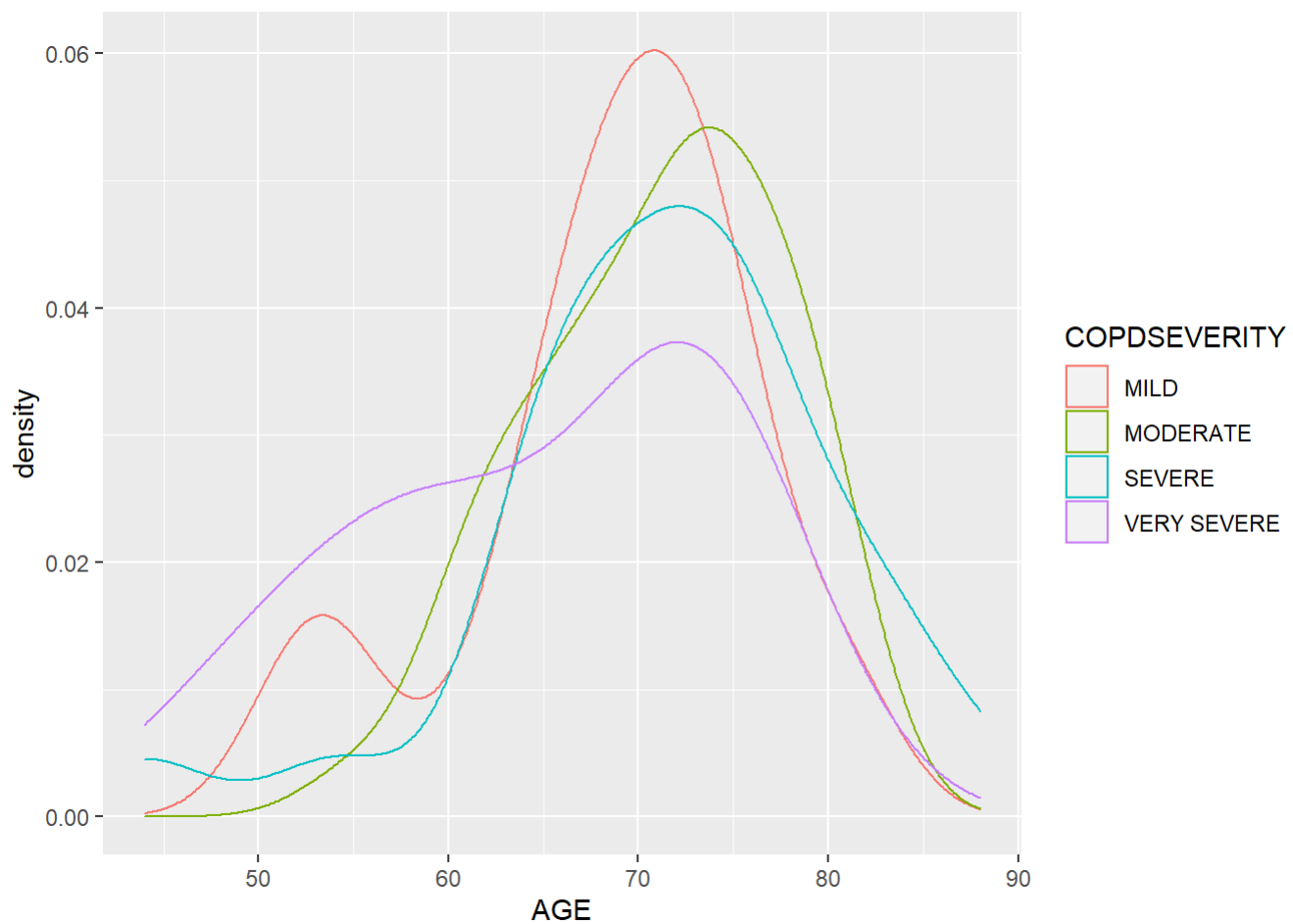
*#Barplot of Age*

```
boxplot(COPD$AGE, horizontal = TRUE, main="Age Distribution forf COPD Dataset", xlab="Years")
```

## Age Distribution for COPD Dataset



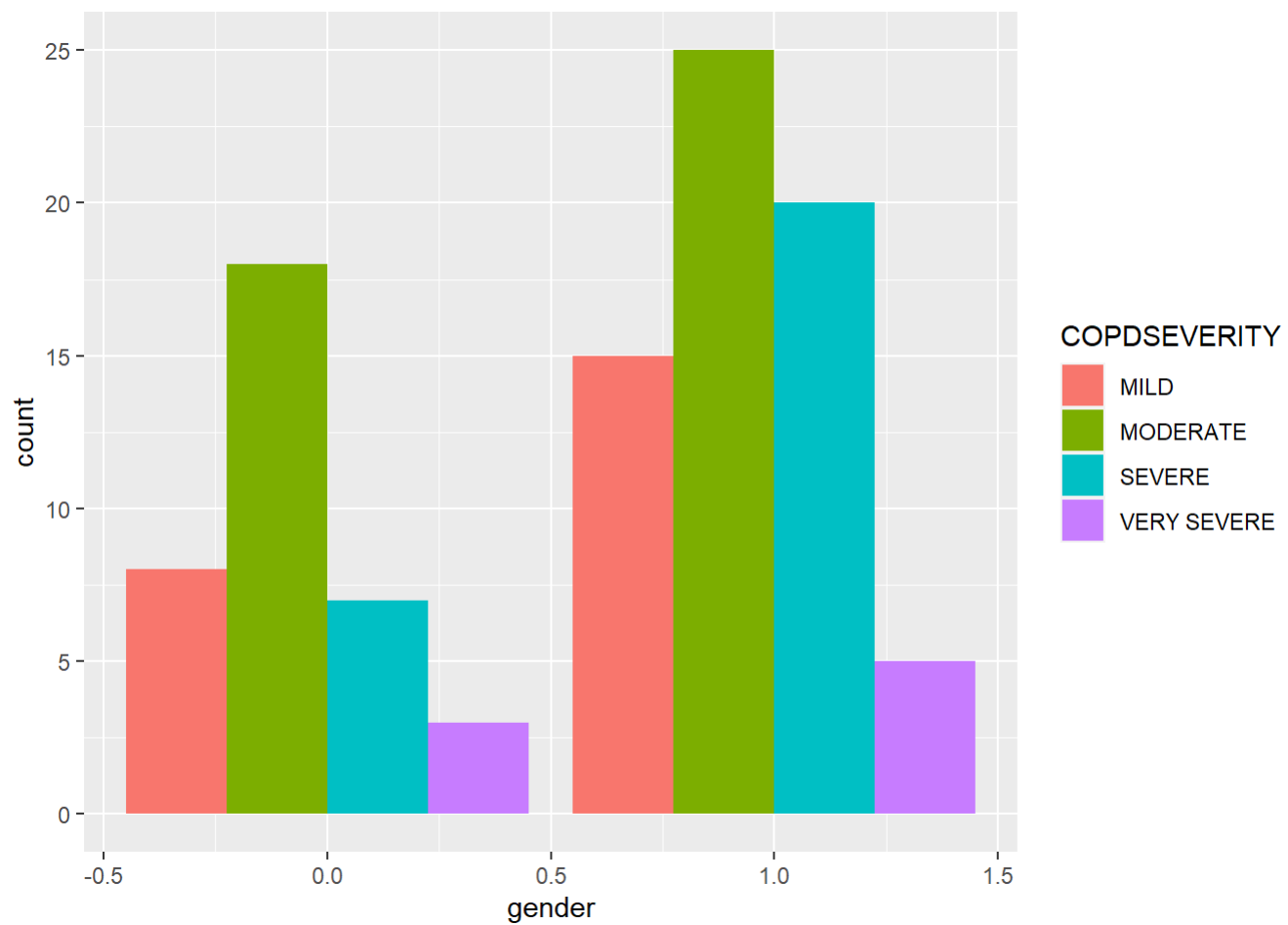
```
#Age by severity  
Age <- ggplot(COPD, aes(x=AGE, col=COPDSEVERITY)) + geom_density()  
Age
```



*#Severity by Gender*

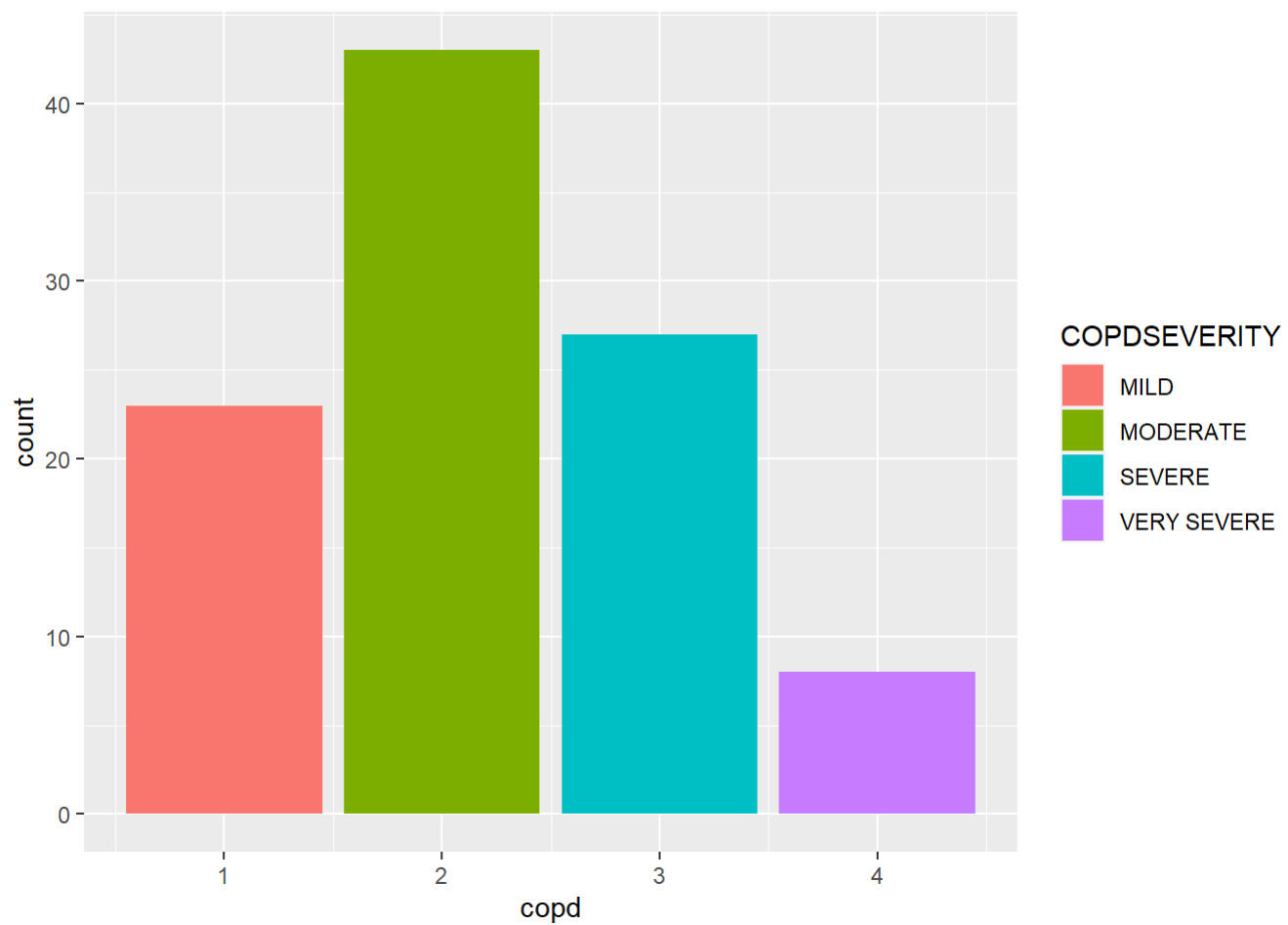
```
ggplot(COPD, aes(x=gender, fill=COPDSEVERITY)) + geom_bar(position="dodge")
```





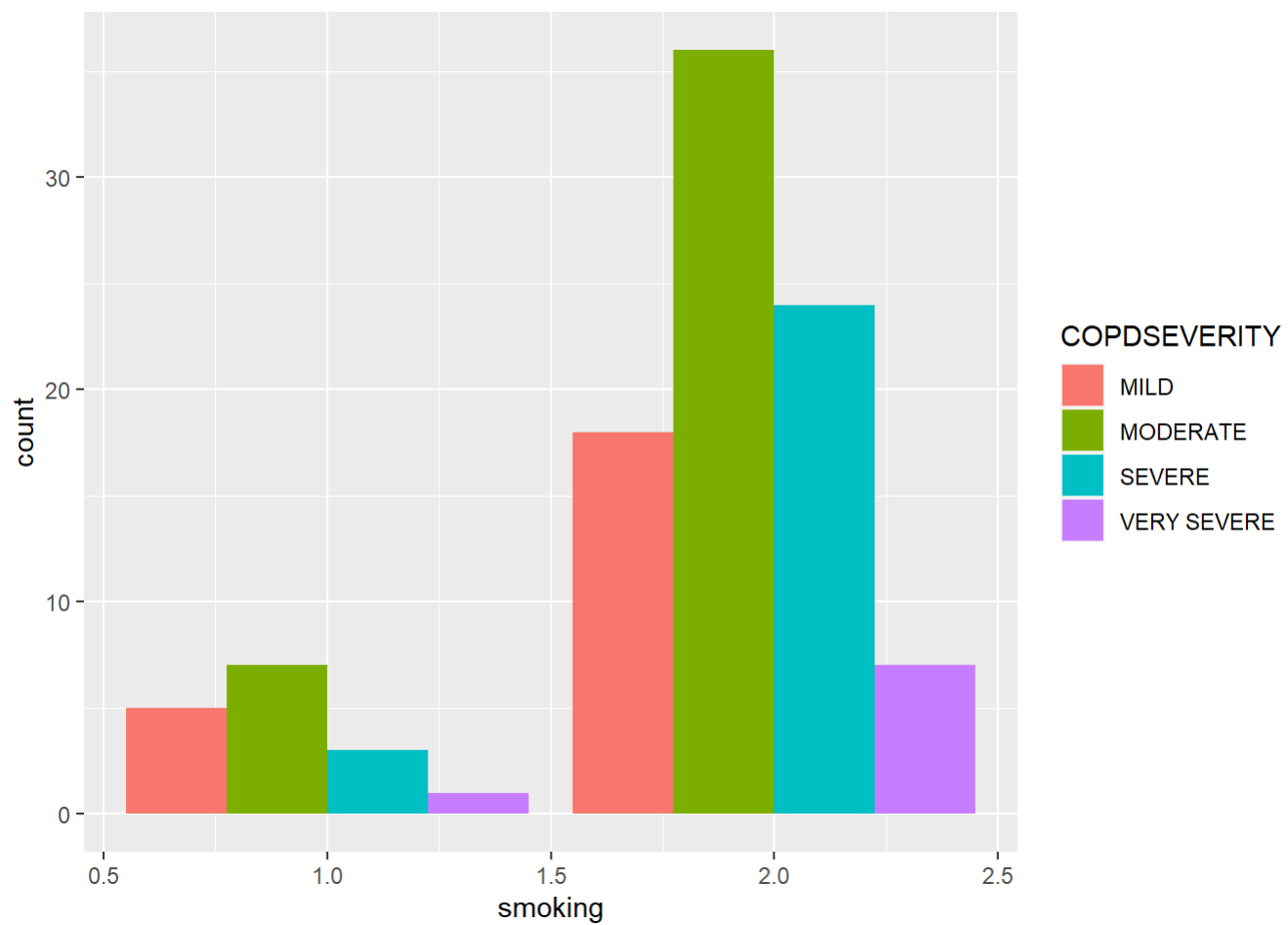
*#Severity by COPD*

```
ggplot(COPD, aes(x=copd, fill=COPDSEVERITY)) + geom_bar(position="dodge")
```



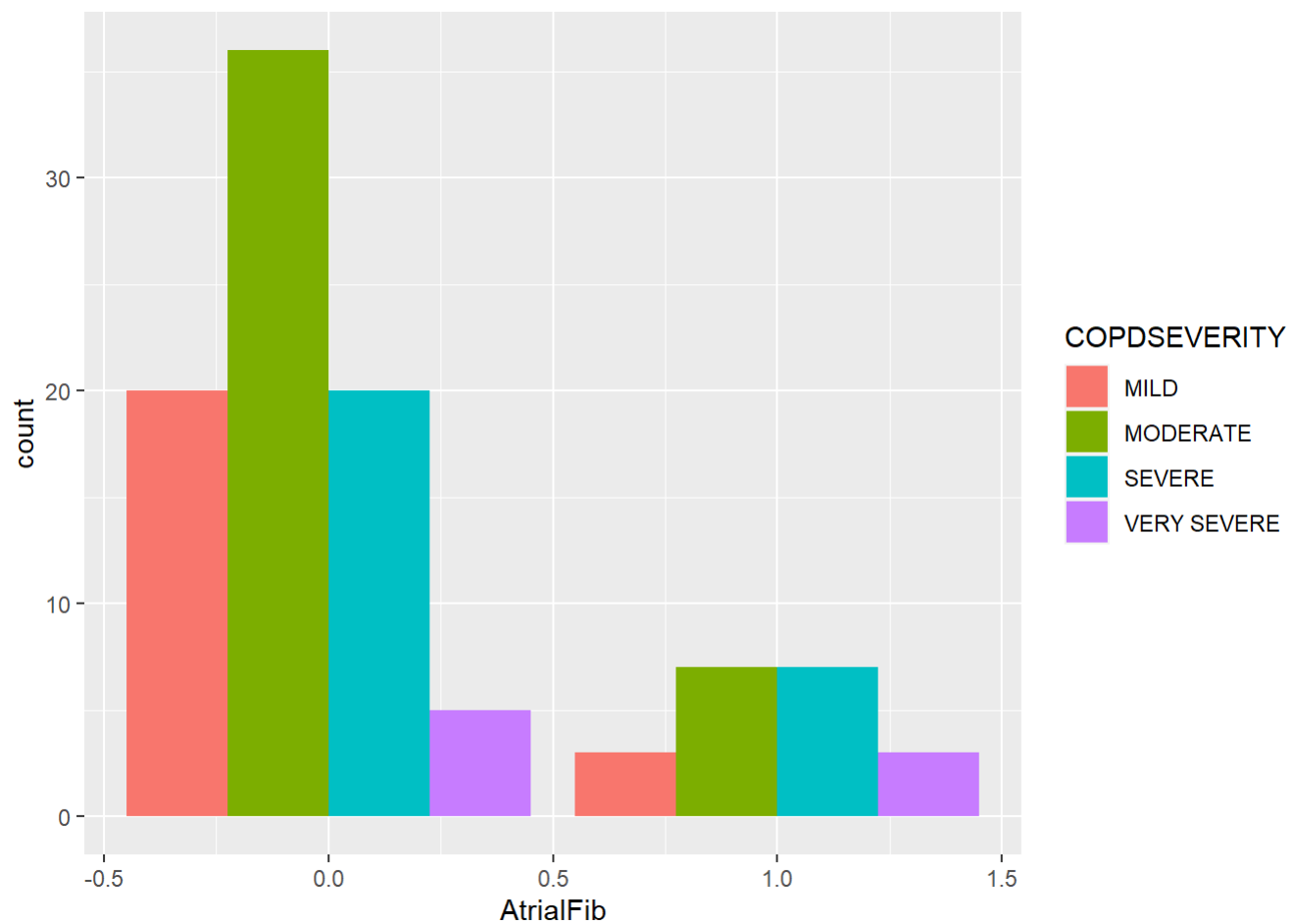
*#Severity by smoking*

```
ggplot(COPD, aes(x=smoking, fill=COPDSEVERITY)) + geom_bar(position="dodge")
```



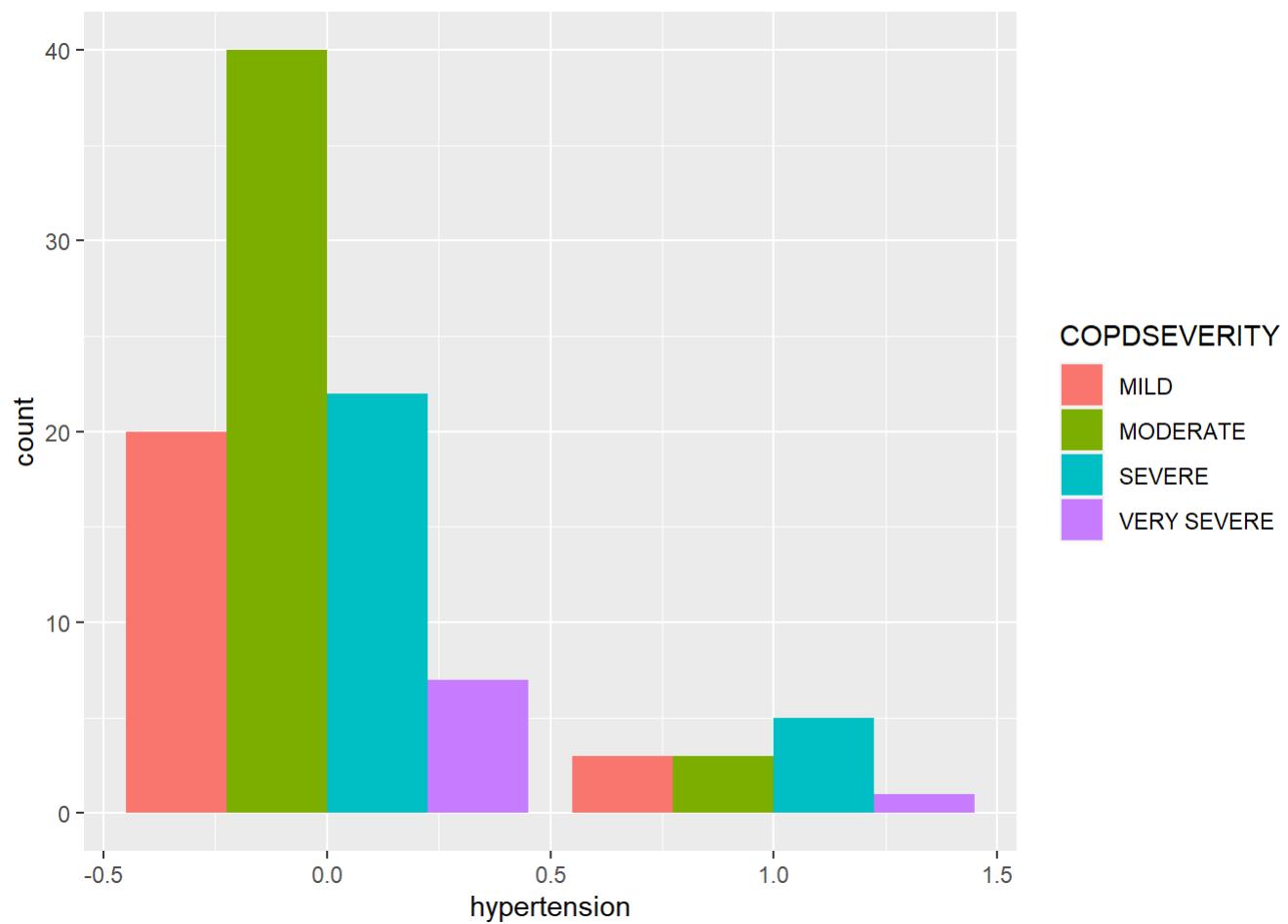
*#Severity by AFib*

```
ggplot(COPD, aes(x=AtrialFib, fill=COPDSEVERITY)) + geom_bar(position="dodge")
```



*#Severity by AFib*

```
ggplot(COPD, aes(x=hypertension, fill=COPDSEVERITY)) + geom_bar(position="dodge")
```



```
#Logistic Regression
```

```
mylogit <- glm(copd ~ smoking + gender + Diabetes + AtrialFib + hypertension, data = COPD)
```

```
summary(mylogit)
```

```
##
## Call:
## glm(formula = copd ~ smoking + gender + Diabetes + AtrialFib +
##      hypertension, data = COPD)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.5513  -0.5110  -0.1699   0.7186   2.0666
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.59716     0.47156   3.387  0.00103 **
## smoking       0.23647     0.24495   0.965  0.33681
## gender        0.09980     0.18695   0.534  0.59470
## Diabetes      0.05959     0.23389   0.255  0.79945
## AtrialFib     0.38136     0.23799   1.602  0.11238
## hypertension  0.11155     0.27998   0.398  0.69122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.7834128)
##
##      Null deviance: 78.040  on 100  degrees of freedom
## Residual deviance: 74.424  on  95  degrees of freedom
## AIC: 269.79
##
## Number of Fisher Scoring iterations: 2
```

```
## CIs using standard errors
confint.default(mylogit)
```

```
##              2.5 %    97.5 %
## (Intercept)  0.67290634 2.5214038
## smoking     -0.24362077 0.7165545
## gender      -0.26662024 0.4662297
## Diabetes    -0.39882625 0.5180039
## AtrialFib   -0.08508844 0.8478140
## hypertension -0.43720075 0.6602962
```

```
## odds ratios and 95% CI
exp(cbind(OR = coef(mylogit), confint(mylogit)))
```

```
## Waiting for profiling to be done...
```

##	OR	2.5 %	97.5 %
## (Intercept)	4.938961	1.9599253	12.446056
## smoking	1.266766	0.7837848	2.047367
## gender	1.104955	0.7659639	1.593973
## Diabetes	1.061400	0.6711073	1.678674
## AtrialFib	1.464279	0.9184311	2.334538
## hypertension	1.118007	0.6458418	1.935366