QBS121_FINAL_PROJECT

Group 3 (Ayush Chakraborty, Gibran Erlangga, Hyunjoe Yoo, Parinitha Kompala, Shrey Khetrapal)

2/26/2022

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5          v purrr   0.3.4
## v tibble  3.1.5          v dplyr   1.0.7.9000
## v tidyr   1.1.4          v stringr 1.4.0
## v readr   2.0.2          v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.1.1

## Warning: package 'tibble' was built under R version 4.1.1

## Warning: package 'tidyr' was built under R version 4.1.1

## Warning: package 'readr' was built under R version 4.1.1

## Warning: package 'stringr' was built under R version 4.1.1

## Warning: package 'forcats' was built under R version 4.1.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
#reading the dataset
insurance<-read.csv("insurance.csv")
head(insurance)
```

```
##   age    sex    bmi children smoker    region   charges
## 1  19 female 27.900        0    yes southwest 16884.924
## 2  18   male 33.770        1     no southeast  1725.552
## 3  28   male 33.000        3     no southeast  4449.462
## 4  33   male 22.705        0     no northwest 21984.471
## 5  32   male 28.880        0     no northwest  3866.855
## 6  31 female 25.740        0     no southeast  3756.622
```

```
#looking into the variables
str(insurance)
```

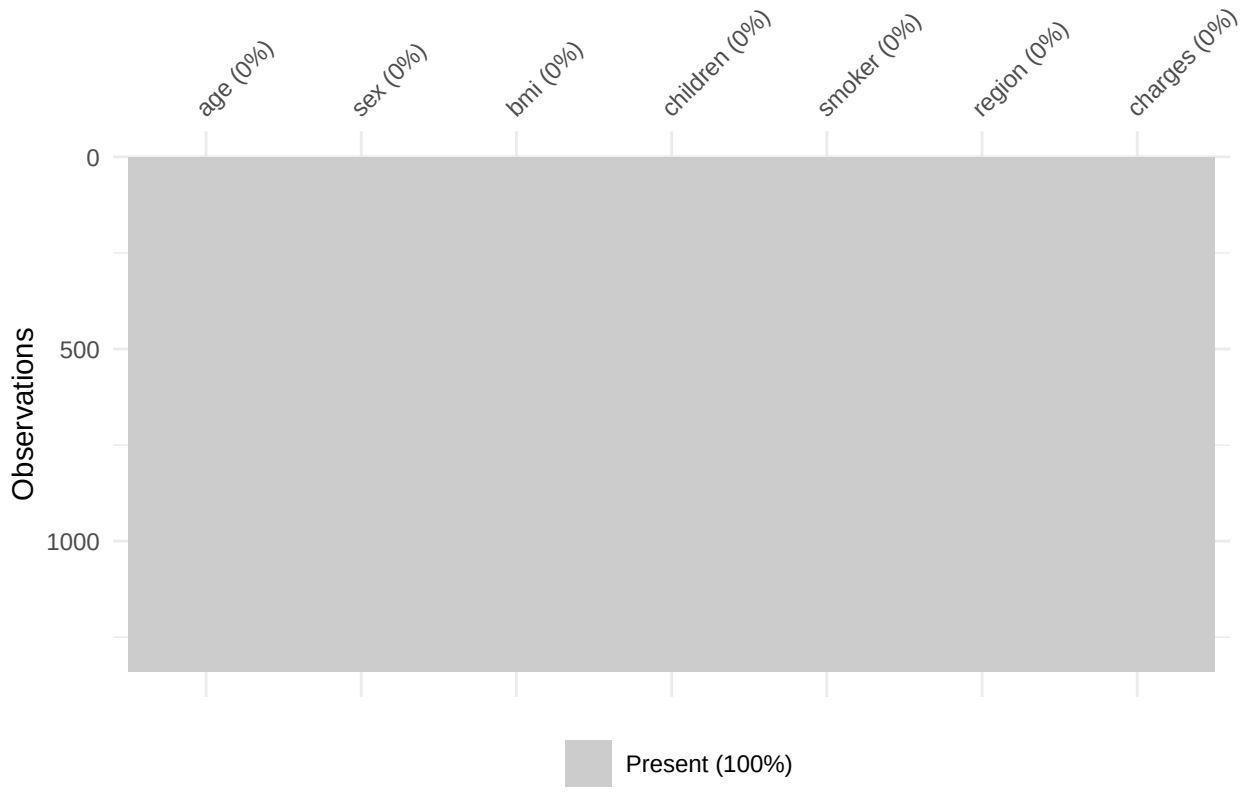
```
## 'data.frame':    1338 obs. of  7 variables:
##  $ age     : int  19 18 28 33 32 31 46 37 37 60 ...
##  $ sex     : chr  "female" "male" "male" "male" ...
##  $ bmi     : num  27.9 33.8 33 22.7 28.9 ...
##  $ children: int  0 1 3 0 0 0 1 3 2 0 ...
##  $ smoker  : chr  "yes" "no" "no" "no" ...
##  $ region  : chr  "southwest" "southeast" "southeast" "northwest" ...
##  $ charges : num  16885 1726 4449 21984 3867 ...
```

```
#converting the str into factor types
insurance$smoker<-as.factor(insurance$smoker)
insurance$sex<-as.factor(insurance$sex)
insurance$region<-as.factor(insurance$region)
Data <- insurance
```

```
#checking for missingness in the dataset.
library(visdat)
```

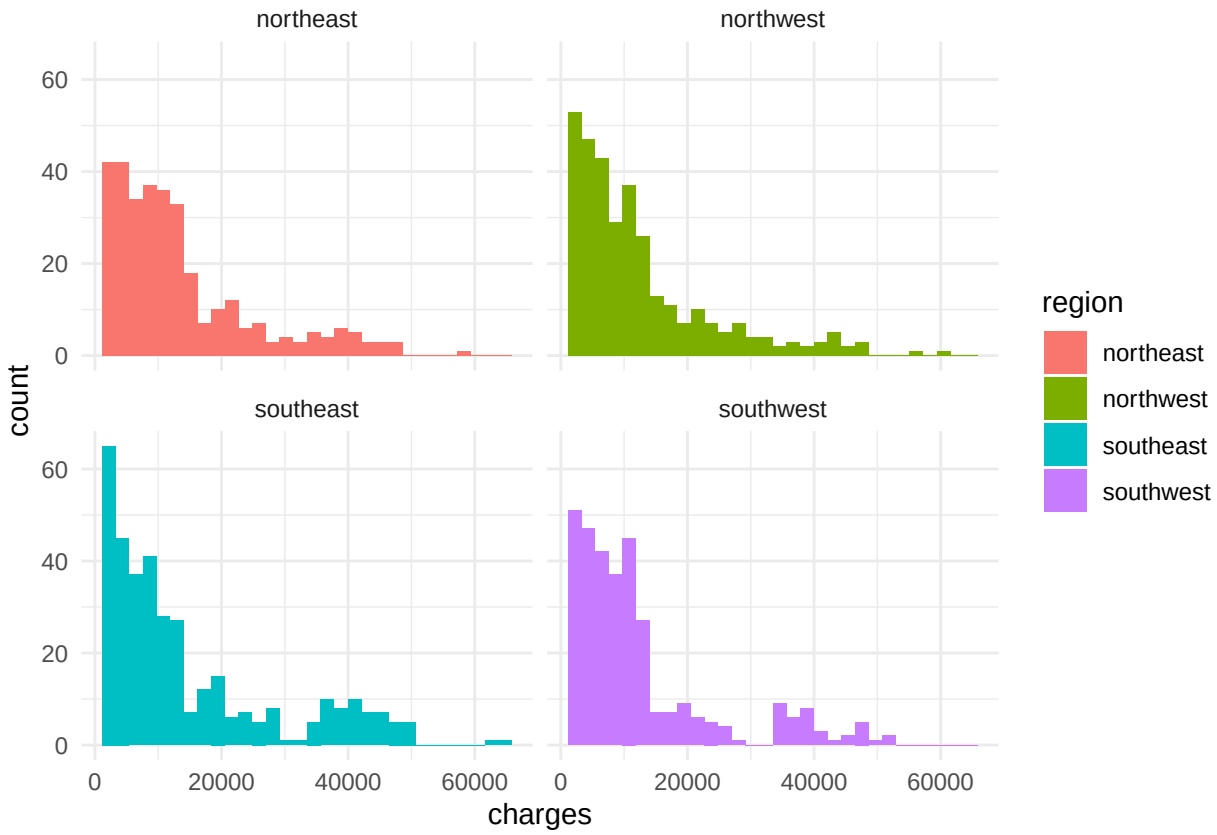
```
## Warning: package 'visdat' was built under R version 4.1.1
```

```
Data %>%
  visdat::vis_miss()
```



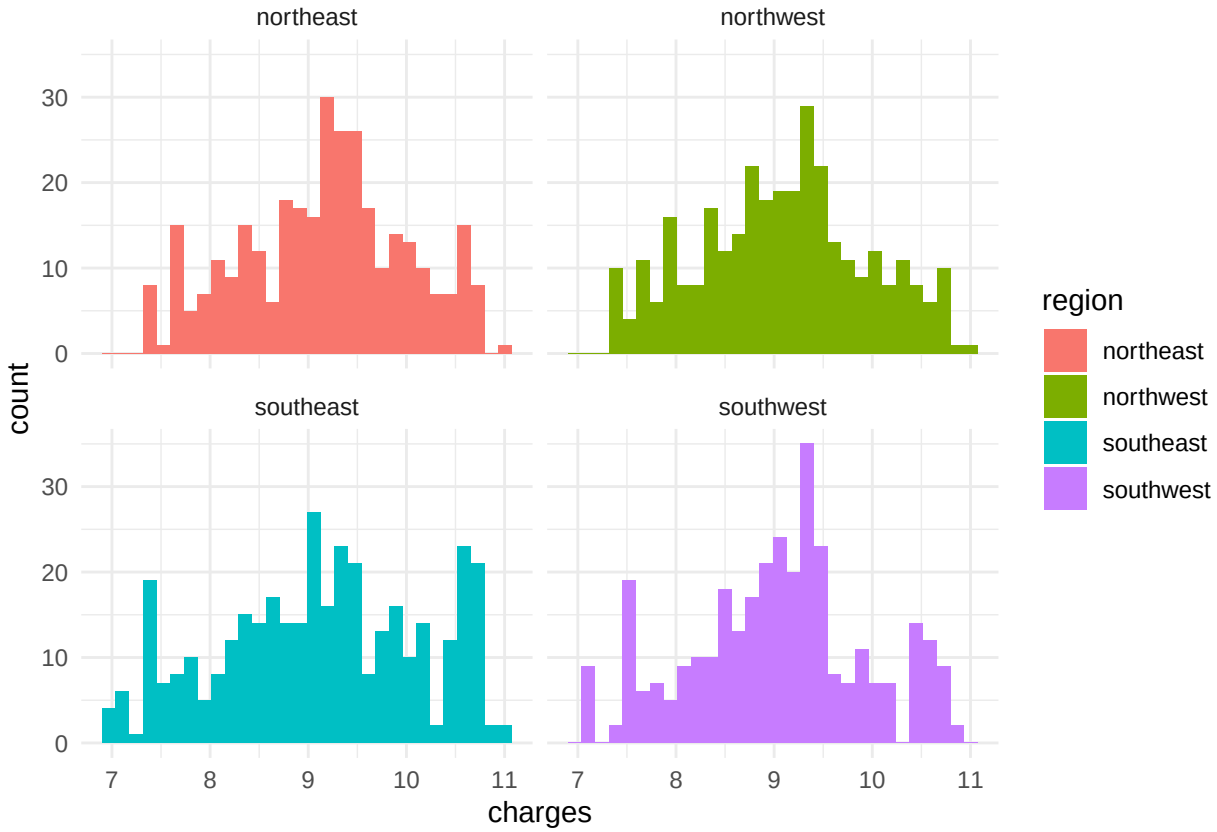
```
#we are plotting the dependent variable
#from this we can concluse that it is not normally distributed
Data %>%
  as_tibble() %>%
  select(region, charges) %>%
  ggplot(aes(charges, fill = region)) +
  geom_histogram() +
  facet_wrap(~region) +
  theme(legend.position = "none") +
  theme_minimal()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
Data %>%
  as_tibble() %>%
  mutate(charges = log(charges)) %>%
  select(region, charges) %>%
  ggplot(aes(charges, fill = region)) +
  geom_histogram() +
  facet_wrap(~region) +
  theme(legend.position = "none") +
  theme_minimal()# to make the linear regression result reliable, the target var should be normally dis
```

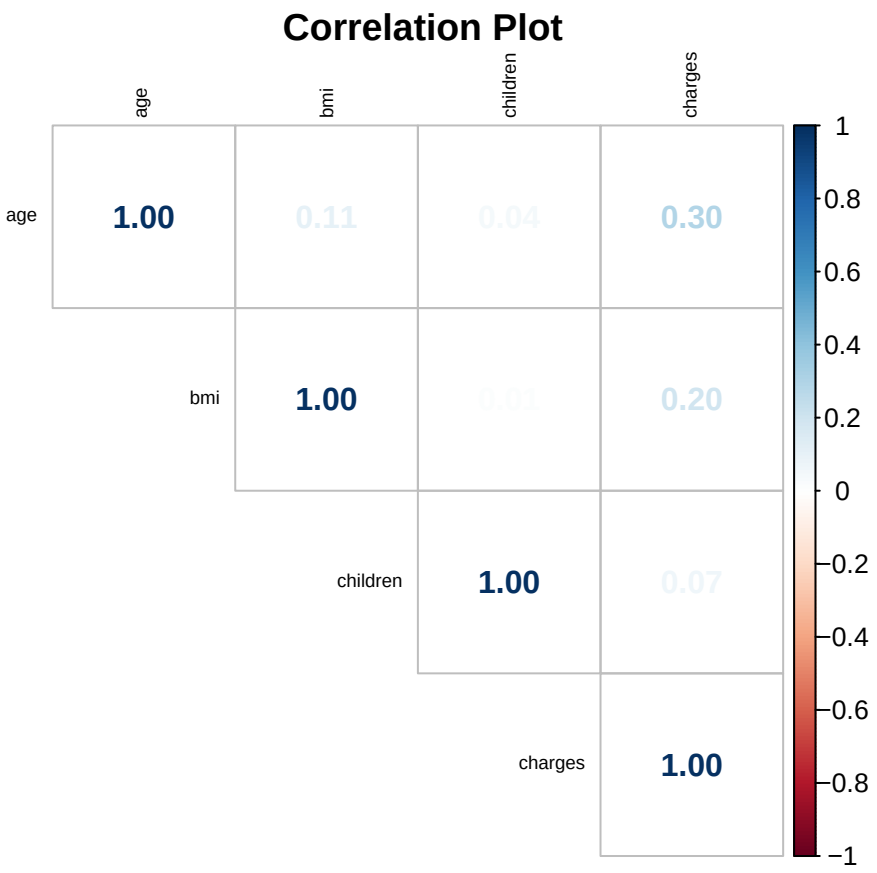
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```




```
#checking corelation between the variables
cor(insurance$charges,insurance$age)#0.2990082
```

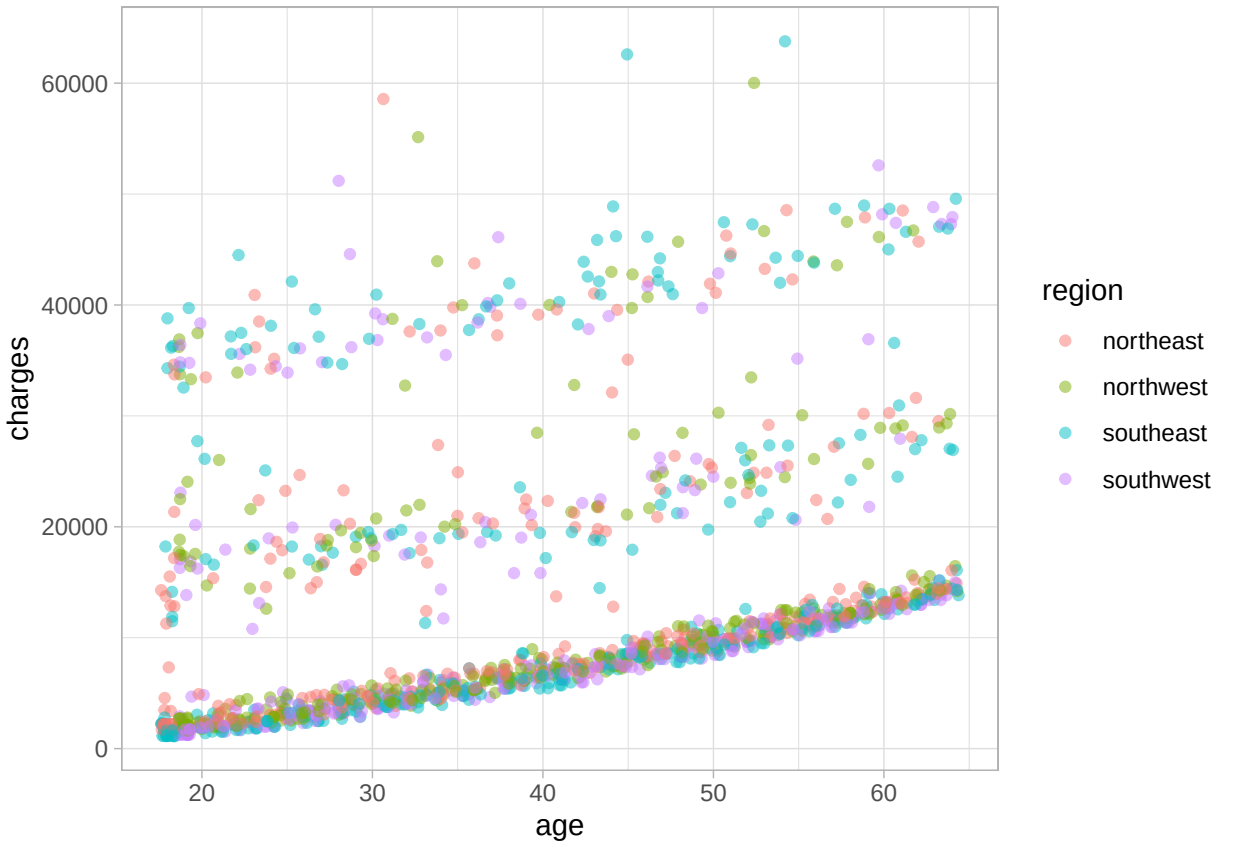
```
## [1] 0.2990082
```

```
insurance_n <- select_if(insurance, is.numeric)
corrmatrix <- cor(insurance_n)
corrplot::corrplot(corrmatrix, method=c("number"), type = "upper",tl.cex=.6
, tl.col="black", title="Correlation Plot",number.font = 2, mar=c(0,0,1,0), )
```

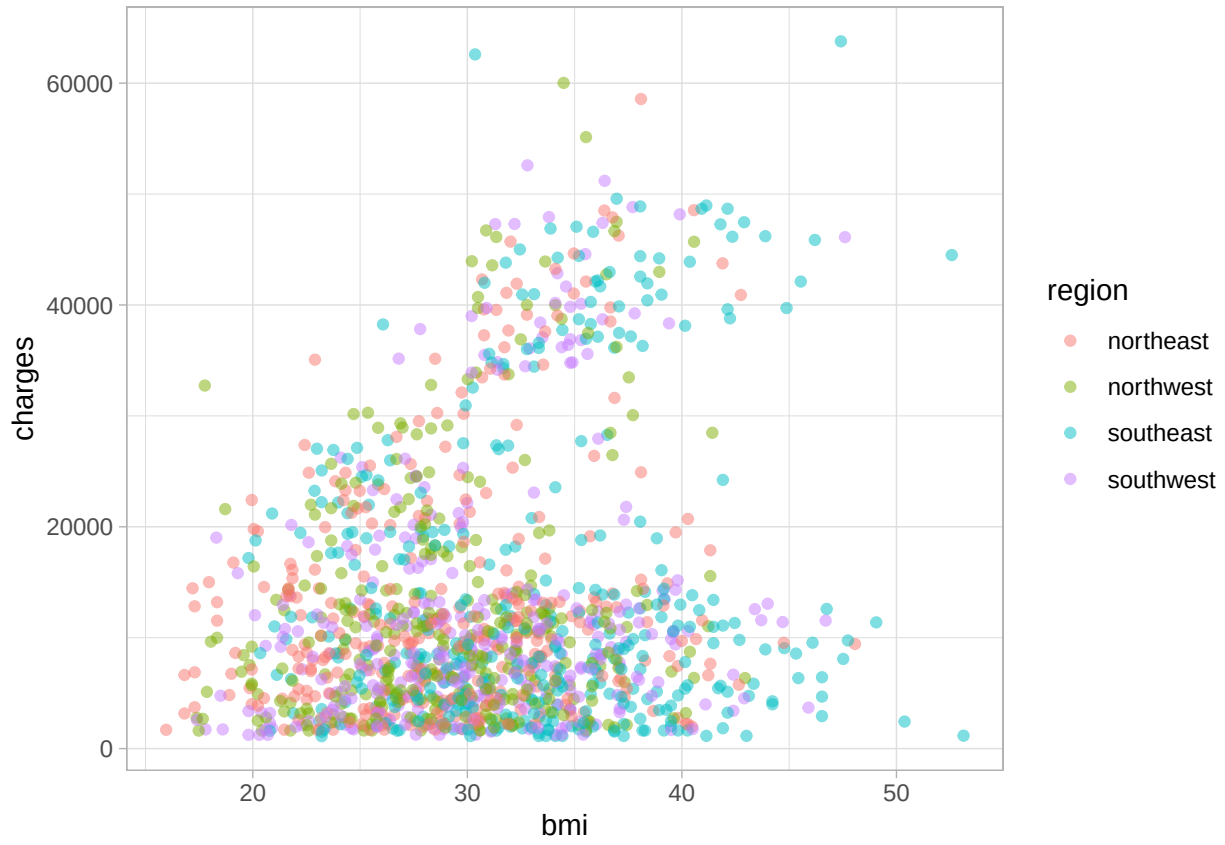



```
#from this we can say that charges comparitively highly corelated with "age" followed by "bmi". And the
```

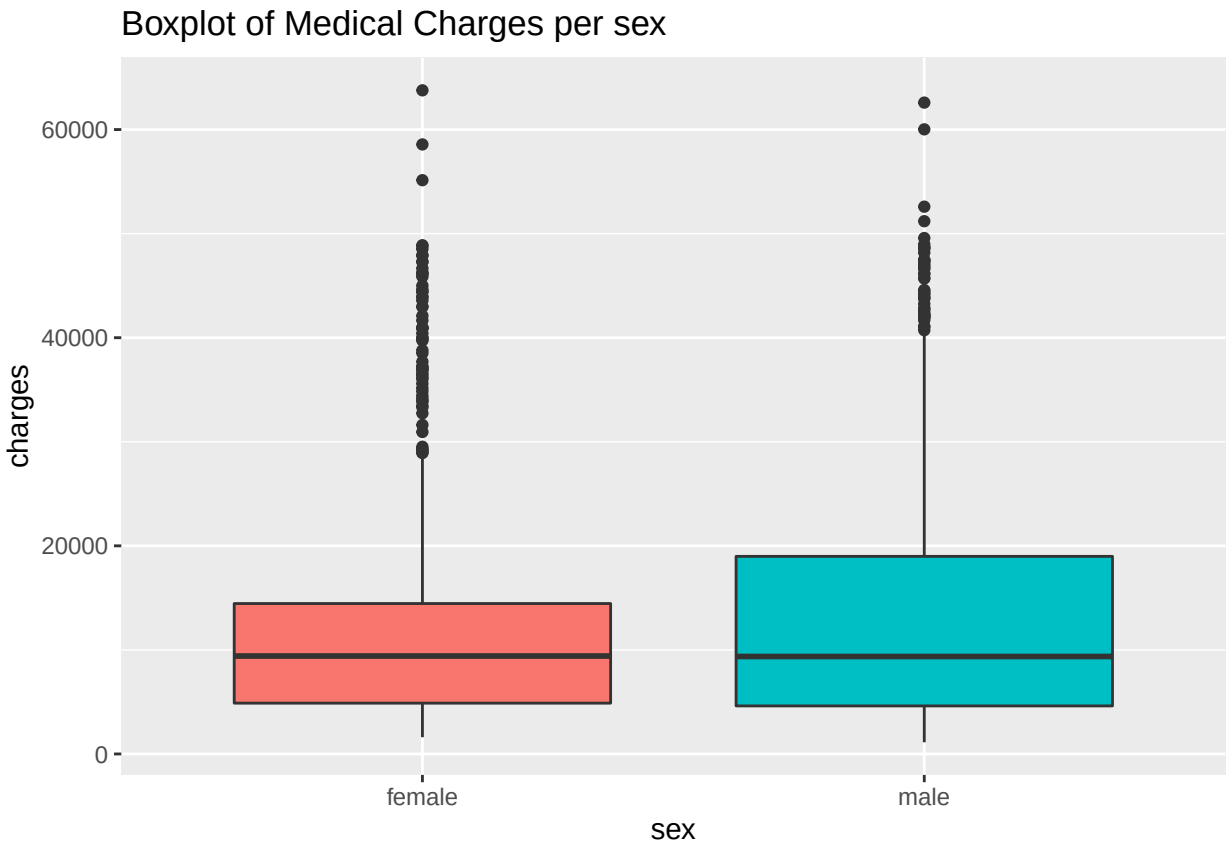
```
#looking into the correlation between charges and age, charges and bmi based on the region
Data<-insurance
x <- ggplot(Data, aes(age, charges,color=region)) +
  geom_jitter( alpha = 0.5) +
    theme_light()

y <- ggplot(Data, aes(bmi, charges, color = region)) +
  geom_jitter( alpha = 0.5) +
  theme_light()

#As Age go up Charges for health insurance also trends up.
x
```

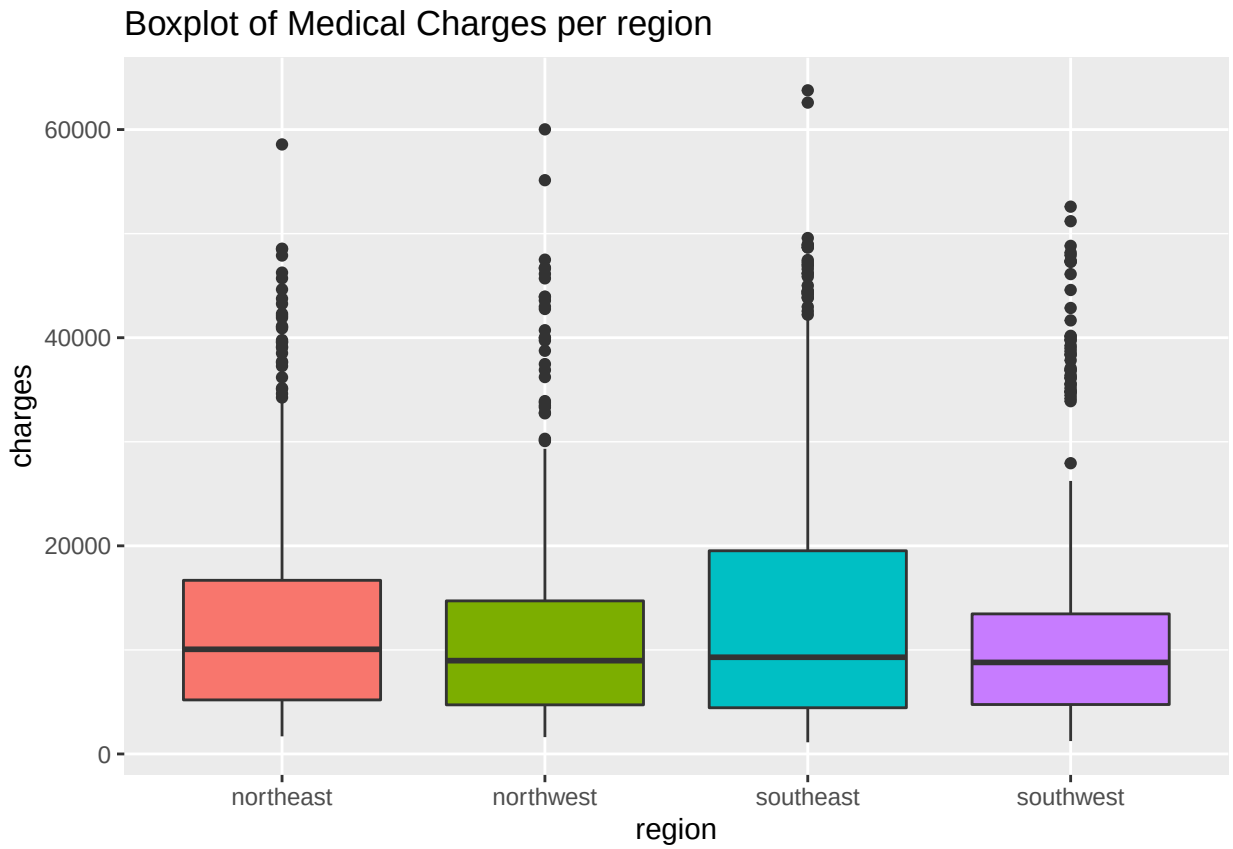

60000
40000
20000
0
charges
20
30
40
50
60
age
region
northeast
northwest
southeast
southwest

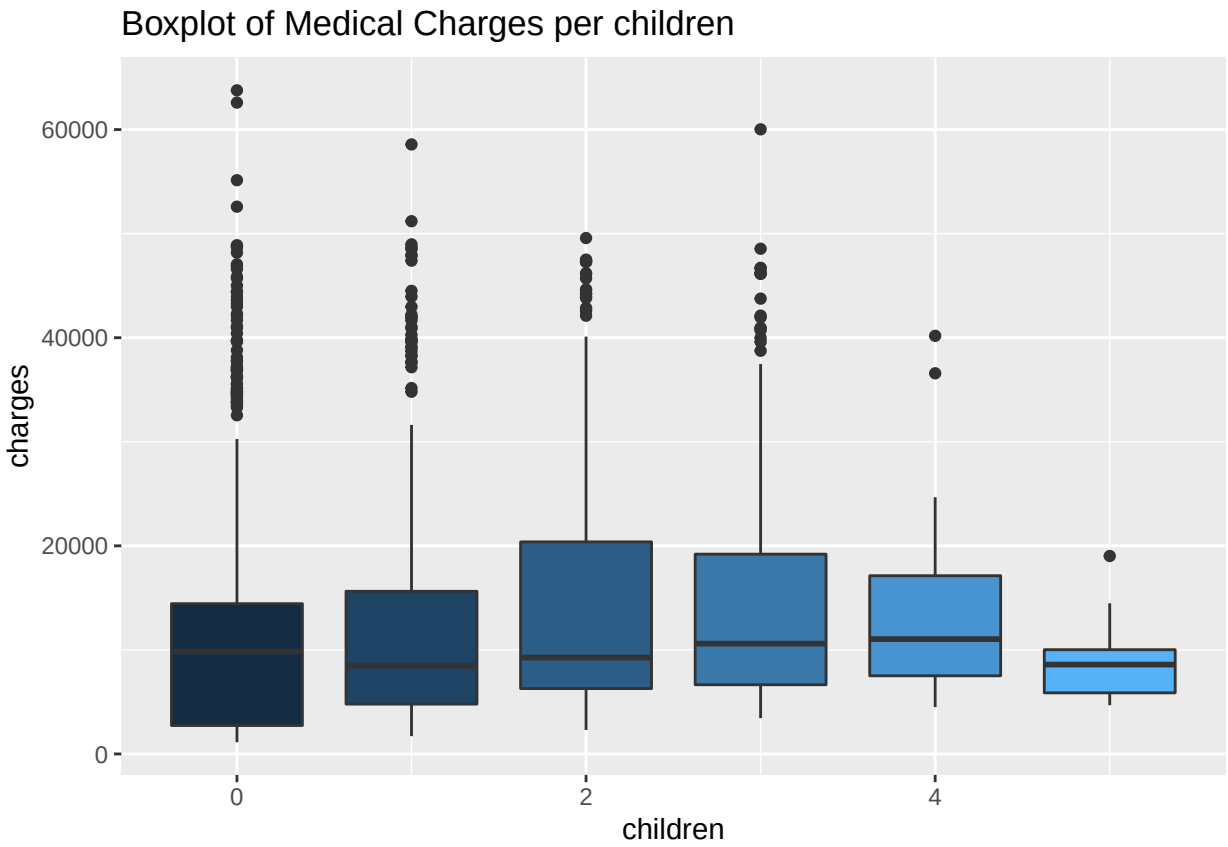
```
y
```

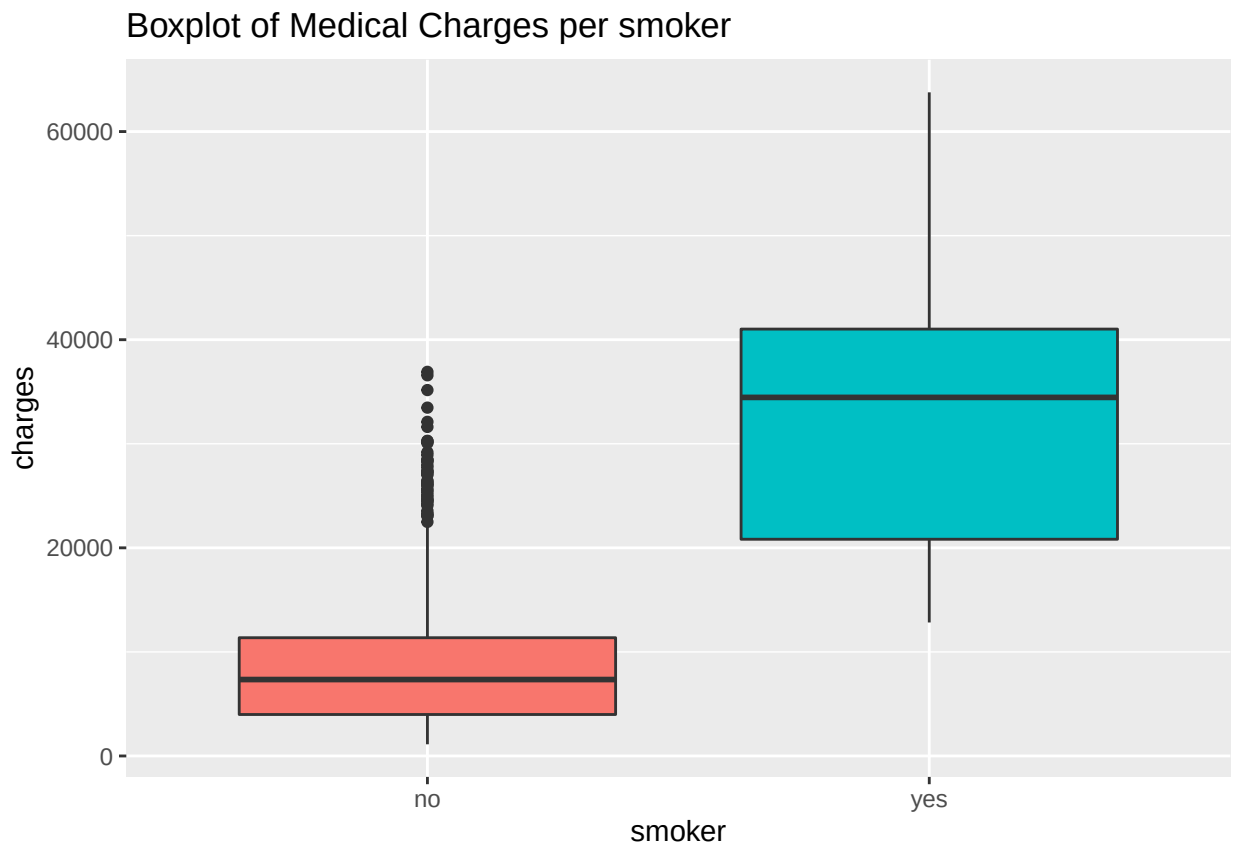



```
#plot between the dependent variable and the all the other independent variable
for (col in c('sex', 'region', 'children', 'smoker')) {
  plot <- ggplot(data = insurance,
                 aes_string(x = col, y = 'charges', group = col, fill = col)) +
            geom_boxplot(show.legend = FALSE) +
            ggtitle(glue::glue("Boxplot of Medical Charges per {col}"))
  print(plot)
}
```


Boxplot of Medical Charges per sex
charges
60000
40000
20000
0
female
male
sex


Boxplot of Medical Charges per region
charges
60000
40000
20000
0
northeast
northwest
southeast
southwest
region


Boxplot of Medical Charges per children
charges
60000
40000
20000
0
0
2
4
children



LMM

```
model1_linear<-lm(charges~age+bmi+region+children+sex+smoker,data=insurance)
summary(model1_linear)
```

```
##
## Call:
## lm(formula = charges ~ age + bmi + region + children + sex +
##     smoker, data = insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11304.9  -2848.1   -982.1   1393.9  29992.8
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -11938.5      987.8 -12.086  < 2e-16 ***
## age                256.9       11.9  21.587  < 2e-16 ***
## bmi                339.2       28.6  11.860  < 2e-16 ***
## regionnorthwest   -353.0      476.3  -0.741 0.458769
## regionsoutheast  -1035.0      478.7  -2.162 0.030782 *
## regionsouthwest   -960.0      477.9  -2.009 0.044765 *
## children           475.5      137.8   3.451 0.000577 ***
```

```
## sexmale           -131.3      332.9  -0.394 0.693348
## smokeryes        23848.5      413.1  57.723  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6062 on 1329 degrees of freedom
## Multiple R-squared:  0.7509, Adjusted R-squared:  0.7494
## F-statistic: 500.8 on 8 and 1329 DF,  p-value: < 2.2e-16
```

```
library(sjPlot)
```

```
## Warning: package 'sjPlot' was built under R version 4.1.1
```

```
## #refugeeswelcome
```

```
library(sjlabelled)
```

```
##
## Attaching package: 'sjlabelled'
```

```
## The following object is masked from 'package:forcats':
##
##     as_factor
```

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

```
## The following object is masked from 'package:ggplot2':
##
##     as_label
```

```
library(sjmisc)
```

```
## Warning: package 'sjmisc' was built under R version 4.1.1
```

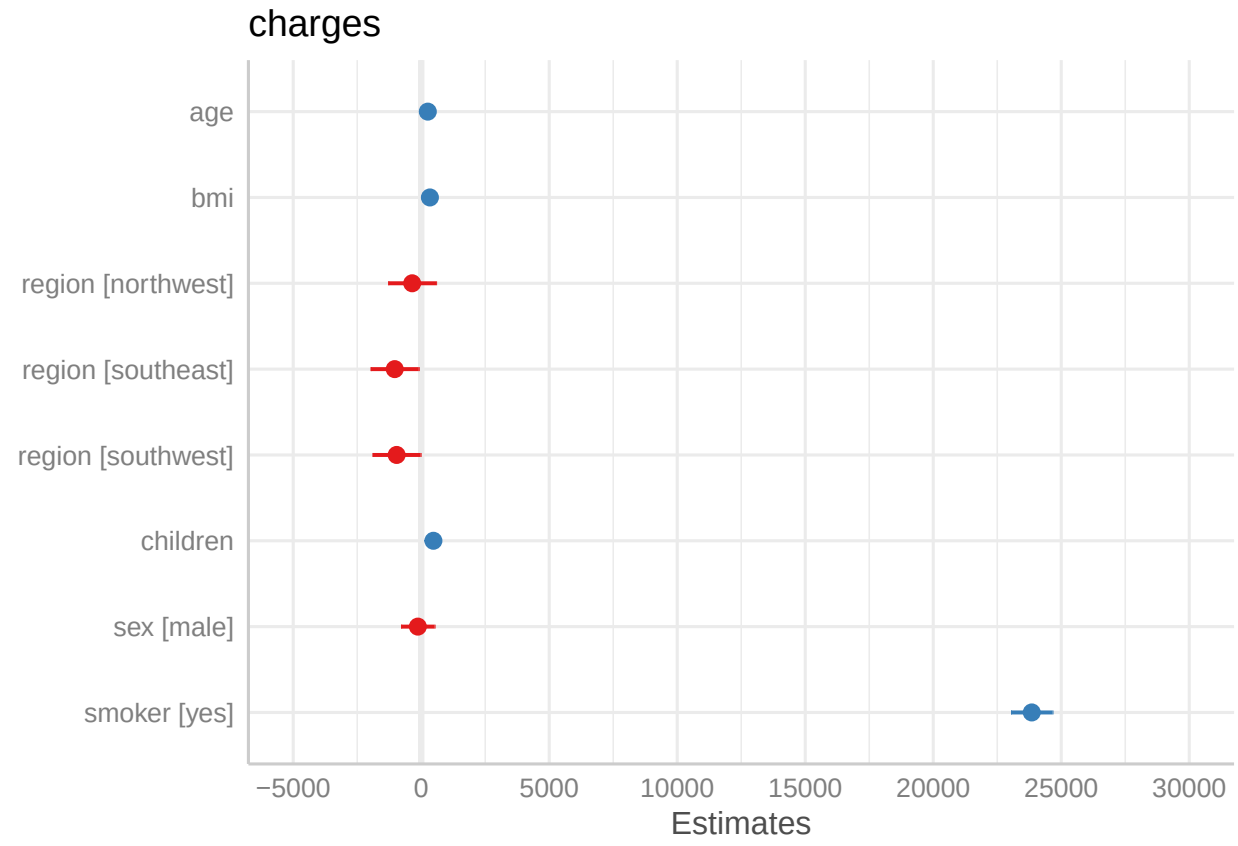
```
##
## Attaching package: 'sjmisc'
```

```
## The following object is masked from 'package:purrr':
##
##     is_empty
```

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

```
## The following object is masked from 'package:tibble':
##
##     add_case
```

```
theme_set(theme_sjplot())
plot_model(model1_linear)
```




```
model2_linear<-lm(charges~age+bmi+smoker,data=insurance)
summary(model2_linear)
```

```
##
## Call:
## lm(formula = charges ~ age + bmi + smoker, data = insurance)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -12415.4  -2970.9   -980.5   1480.0  28971.8
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -11676.83     937.57  -12.45   <2e-16 ***
## age            259.55      11.93   21.75   <2e-16 ***
## bmi            322.62      27.49   11.74   <2e-16 ***
## smokeryes    23823.68     412.87   57.70   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6092 on 1334 degrees of freedom
## Multiple R-squared:  0.7475, Adjusted R-squared:  0.7469
```

```
## F-statistic:  1316 on 3 and 1334 DF,  p-value: < 2.2e-16
```

GLM

```
# We want to classify patient as Risky or Safe based on some conditions. Looking at the dataset we can
# We use BMI > 25, age > 50 and smoker status = YES to categories our patient
insurance$risk <- ifelse(insurance$bmi > 25 & insurance$age > 50 & insurance$smoker == "yes", 1, 0)

# Down sampling data with 0 to match the 1
insurance %>%
  filter(risk == 0) %>%
  slice_sample(n = 80) -> insurance_risk_0

insurance %>%
  filter(risk == 1) -> insurance_risk_1

insurance_risk_0 %>%
  rbind(insurance_risk_1) -> downsampled_data

# Now splitting our dataset into test and train

library(tidymodels)
```

```
## Warning: package 'tidymodels' was built under R version 4.1.1
```

```
## Registered S3 method overwritten by 'tune':
##   method                   from
##   required_pkgs.model_spec parsnip
```

```
## -- Attaching packages --------------------------------------- tidymodels 0.1.4 --
```

```
## v broom        0.7.10     v rsample      0.1.1
## v dials        0.1.0      v tune         0.1.6
## v infer        1.0.0      v workflows    0.2.4
## v modeldata    0.1.1      v workflowsets 0.1.0
## v parsnip      0.1.7      v yardstick    0.0.9
## v recipes      0.1.17
```

```
## Warning: package 'broom' was built under R version 4.1.1
```

```
## Warning: package 'dials' was built under R version 4.1.1
```

```
## Warning: package 'infer' was built under R version 4.1.1
```

```
## Warning: package 'recipes' was built under R version 4.1.1
```

```
## Warning: package 'rsample' was built under R version 4.1.1
```

```
## Warning: package 'workflows' was built under R version 4.1.1
```

```
## Warning: package 'workflowsets' was built under R version 4.1.1

## Warning: package 'yardstick' was built under R version 4.1.1

## -- Conflicts ---------------------------------------- tidymodels_conflicts() --
## x sjmisc::add_case()         masks tibble::add_case()
## x workflows::add_variables() masks sjmisc::add_variables()
## x sjlabelled::as_label()     masks dplyr::as_label(), ggplot2::as_label()
## x scales::discard()          masks purrr::discard()
## x dplyr::filter()            masks stats::filter()
## x recipes::fixed()           masks stringr::fixed()
## x sjmisc::is_empty()         masks purrr::is_empty()
## x dplyr::lag()               masks stats::lag()
## x sjmisc::replace_na()       masks tidyr::replace_na()
## x yardstick::spec()          masks readr::spec()
## x recipes::step()            masks stats::step()
## * Learn how to get started at https://www.tidymodels.org/start/
```

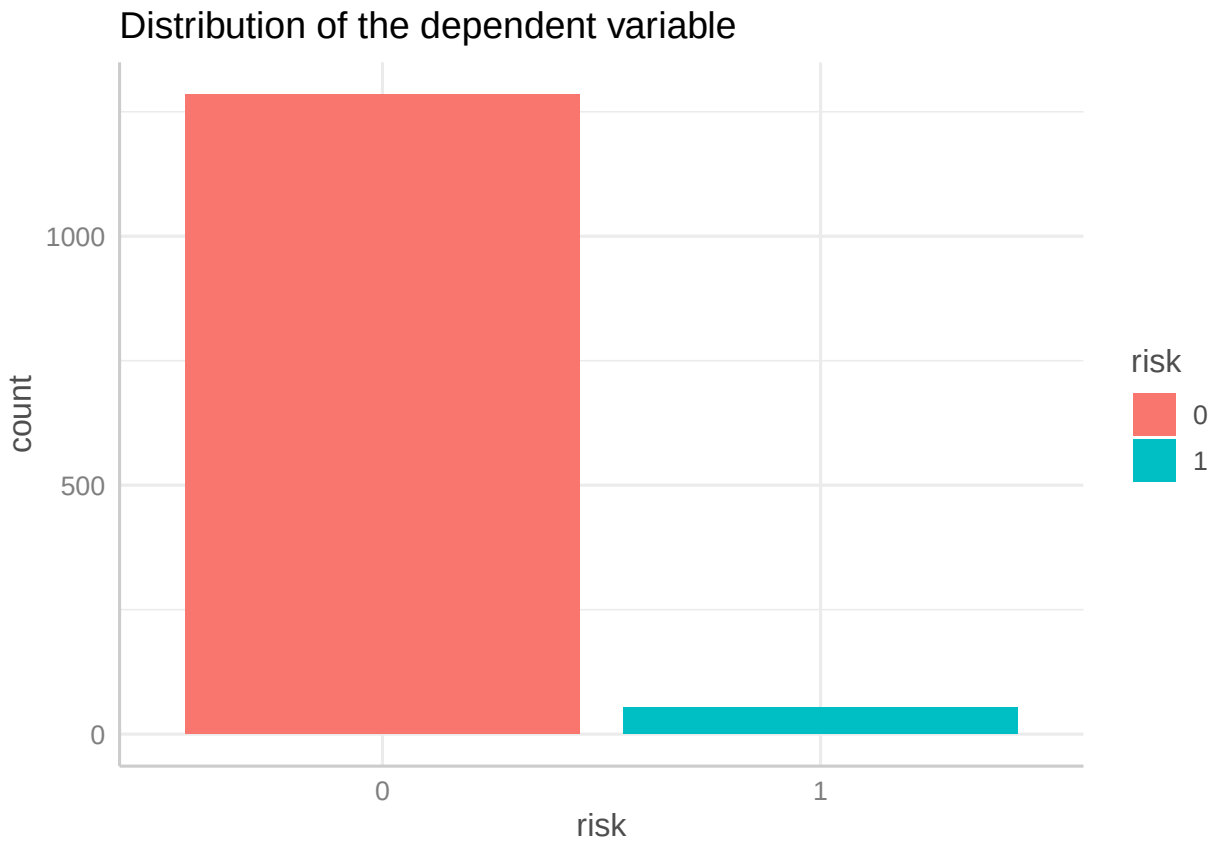
```
insurance_split <- initial_split(downsampled_data)
training_data <- training(insurance_split)
testing_data <- testing(insurance_split)


model_glm <- glm(risk~age+sex+bmi+children+region,
                 family=binomial,
                 data=training_data)
summary(model_glm)
```

```
##
## Call:
## glm(formula = risk ~ age + sex + bmi + children + region, family = binomial,
##     data = training_data)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.30892  -0.39647  -0.02136   0.60233   1.72955
##
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)     -15.17928    4.08687  -3.714 0.000204 ***
## age               0.24322    0.05909   4.116 3.86e-05 ***
## sexmale          -0.15328    0.61084  -0.251 0.801868
## bmi               0.05998    0.05814   1.032 0.302298
## children          0.30526    0.31087   0.982 0.326121
## regionnorthwest  -0.19759    0.89023  -0.222 0.824354
## regionsoutheast   0.04953    0.91549   0.054 0.956858
## regionsouthwest  -1.85316    1.08420  -1.709 0.087405 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 130.861  on 98  degrees of freedom
```

```
## Residual deviance:  69.752  on 91  degrees of freedom
## AIC: 85.752
##
## Number of Fisher Scoring iterations: 7
```

```
insurance %>%
  mutate(risk = risk %>% as.factor()) %>%
  ggplot(aes(risk, fill = risk)) +
  geom_histogram(stat = "count") +
  labs(title = "Distribution of the dependent variable")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```




```
  theme_minimal() +
  theme(legend.position = "none")
```

```
## List of 93
##  $ line                      :List of 6
##   ..$ colour       : chr "black"
##   ..$ size         : num 0.5
##   ..$ linetype     : num 1
##   ..$ lineend      : chr "butt"
##   ..$ arrow        : logi FALSE
##   ..$ inherit.blank: logi TRUE
```

```
##   ..- attr(*, "class")= chr [1:2] "element_line" "element"
##  $ rect                      :List of 5
##   ..$ fill         : chr "white"
##   ..$ colour       : chr "black"
##   ..$ size         : num 0.5
##   ..$ linetype     : num 1
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_rect" "element"
##  $ text                      :List of 11
##   ..$ family       : chr ""
##   ..$ face         : chr "plain"
##   ..$ colour       : chr "black"
##   ..$ size         : num 11
##   ..$ hjust        : num 0.5
##   ..$ vjust        : num 0.5
##   ..$ angle        : num 0
##   ..$ lineheight   : num 0.9
##   ..$ margin       : 'margin' num [1:4] 0points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ title                     : NULL
##  $ aspect.ratio              : NULL
##  $ axis.title                : NULL
##  $ axis.title.x              :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 1
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 2.75points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.title.x.top          :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 0
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 2.75points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.title.x.bottom       : NULL
```

```
##  $ axis.title.y              :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 1
##   ..$ angle        : num 90
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 2.75points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.title.y.left         : NULL
##  $ axis.title.y.right        :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 0
##   ..$ angle        : num -90
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 0points 2.75points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text                 :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : chr "grey30"
##   ..$ size         : 'rel' num 0.8
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.x               :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 1
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 2.2points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
```

```
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.x.top           :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : num 0
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 2.2points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.x.bottom        : NULL
##  $ axis.text.y               :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 1
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 2.2points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text.y.left          : NULL
##  $ axis.text.y.right         :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 0
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 0points 2.2points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.ticks                : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ axis.ticks.x              : NULL
##  $ axis.ticks.x.top          : NULL
##  $ axis.ticks.x.bottom       : NULL
##  $ axis.ticks.y              : NULL
##  $ axis.ticks.y.left         : NULL
##  $ axis.ticks.y.right        : NULL
```

```
##  $ axis.ticks.length         : 'simpleUnit' num 2.75points
##   ..- attr(*, "unit")= int 8
##  $ axis.ticks.length.x       : NULL
##  $ axis.ticks.length.x.top   : NULL
##  $ axis.ticks.length.x.bottom: NULL
##  $ axis.ticks.length.y       : NULL
##  $ axis.ticks.length.y.left  : NULL
##  $ axis.ticks.length.y.right : NULL
##  $ axis.line                 : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ axis.line.x               : NULL
##  $ axis.line.x.top           : NULL
##  $ axis.line.x.bottom        : NULL
##  $ axis.line.y               : NULL
##  $ axis.line.y.left          : NULL
##  $ axis.line.y.right         : NULL
##  $ legend.background         : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ legend.margin             : 'margin' num [1:4] 5.5points 5.5points 5.5points 5.5points
##   ..- attr(*, "unit")= int 8
##  $ legend.spacing            : 'simpleUnit' num 11points
##   ..- attr(*, "unit")= int 8
##  $ legend.spacing.x          : NULL
##  $ legend.spacing.y          : NULL
##  $ legend.key                : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ legend.key.size           : 'simpleUnit' num 1.2lines
##   ..- attr(*, "unit")= int 3
##  $ legend.key.height         : NULL
##  $ legend.key.width          : NULL
##  $ legend.text               :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : 'rel' num 0.8
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ legend.text.align         : NULL
##  $ legend.title              :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 0
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
```

```
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ legend.title.align        : NULL
##  $ legend.position           : chr "none"
##  $ legend.direction          : NULL
##  $ legend.justification      : chr "center"
##  $ legend.box                : NULL
##  $ legend.box.just           : NULL
##  $ legend.box.margin         : 'margin' num [1:4] 0cm 0cm 0cm 0cm
##   ..- attr(*, "unit")= int 1
##  $ legend.box.background     : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ legend.box.spacing        : 'simpleUnit' num 11points
##   ..- attr(*, "unit")= int 8
##  $ panel.background          : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ panel.border              : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ panel.spacing             : 'simpleUnit' num 5.5points
##   ..- attr(*, "unit")= int 8
##  $ panel.spacing.x           : NULL
##  $ panel.spacing.y           : NULL
##  $ panel.grid                :List of 6
##   ..$ colour       : chr "grey92"
##   ..$ size         : NULL
##   ..$ linetype     : NULL
##   ..$ lineend      : NULL
##   ..$ arrow        : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_line" "element"
##  $ panel.grid.major          : NULL
##  $ panel.grid.minor          :List of 6
##   ..$ colour       : NULL
##   ..$ size         : 'rel' num 0.5
##   ..$ linetype     : NULL
##   ..$ lineend      : NULL
##   ..$ arrow        : logi FALSE
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_line" "element"
##  $ panel.grid.major.x        : NULL
##  $ panel.grid.major.y        : NULL
##  $ panel.grid.minor.x        : NULL
##  $ panel.grid.minor.y        : NULL
##  $ panel.ontop               : logi FALSE
##  $ plot.background           : list()
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ plot.title                :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : 'rel' num 1.2
##   ..$ hjust        : num 0
##   ..$ vjust        : num 1
```

```
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 5.5points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ plot.title.position       : chr "panel"
##  $ plot.subtitle             :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : num 0
##   ..$ vjust        : num 1
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 0points 0points 5.5points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ plot.caption              :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : 'rel' num 0.8
##   ..$ hjust        : num 1
##   ..$ vjust        : num 1
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 5.5points 0points 0points 0points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ plot.caption.position     : chr "panel"
##  $ plot.tag                  :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : 'rel' num 1.2
##   ..$ hjust        : num 0.5
##   ..$ vjust        : num 0.5
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ plot.tag.position         : chr "topleft"
##  $ plot.margin               : 'margin' num [1:4] 5.5points 5.5points 5.5points 5.5points
##   ..- attr(*, "unit")= int 8
##  $ strip.background          : list()
```

```
##   ..- attr(*, "class")= chr [1:2] "element_blank" "element"
##  $ strip.background.x        : NULL
##  $ strip.background.y        : NULL
##  $ strip.placement           : chr "inside"
##  $ strip.text                :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : chr "grey10"
##   ..$ size         : 'rel' num 0.8
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : 'margin' num [1:4] 4.4points 4.4points 4.4points 4.4points
##   .. ..- attr(*, "unit")= int 8
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ strip.text.x              : NULL
##  $ strip.text.y              :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : num -90
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ strip.switch.pad.grid     : 'simpleUnit' num 2.75points
##   ..- attr(*, "unit")= int 8
##  $ strip.switch.pad.wrap     : 'simpleUnit' num 2.75points
##   ..- attr(*, "unit")= int 8
##  $ strip.text.y.left         :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : NULL
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : num 90
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi TRUE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi TRUE
##  - attr(*, "validate")= logi TRUE
```

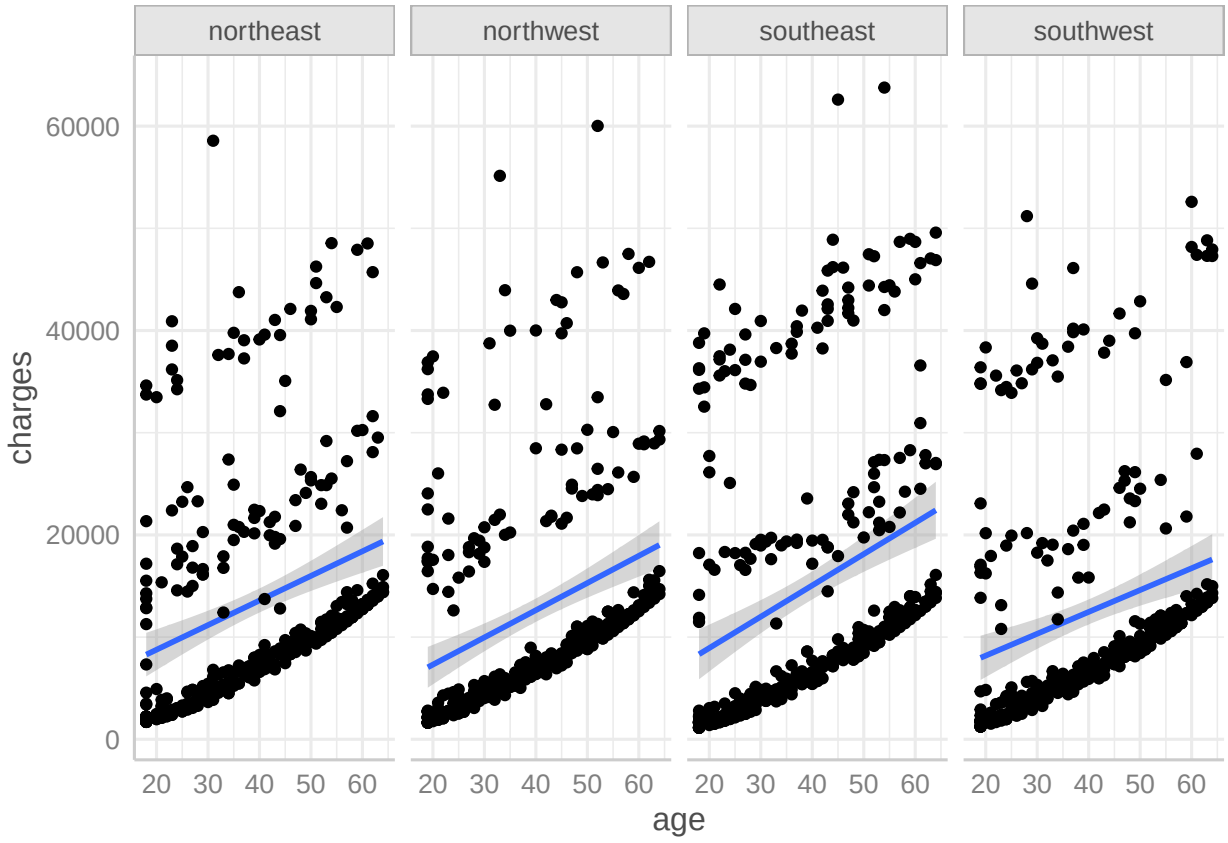
```
# Looking at the number of values in each category
table(insurance$risk)
```

```
##
##    0    1
## 1285   53
```

LMM

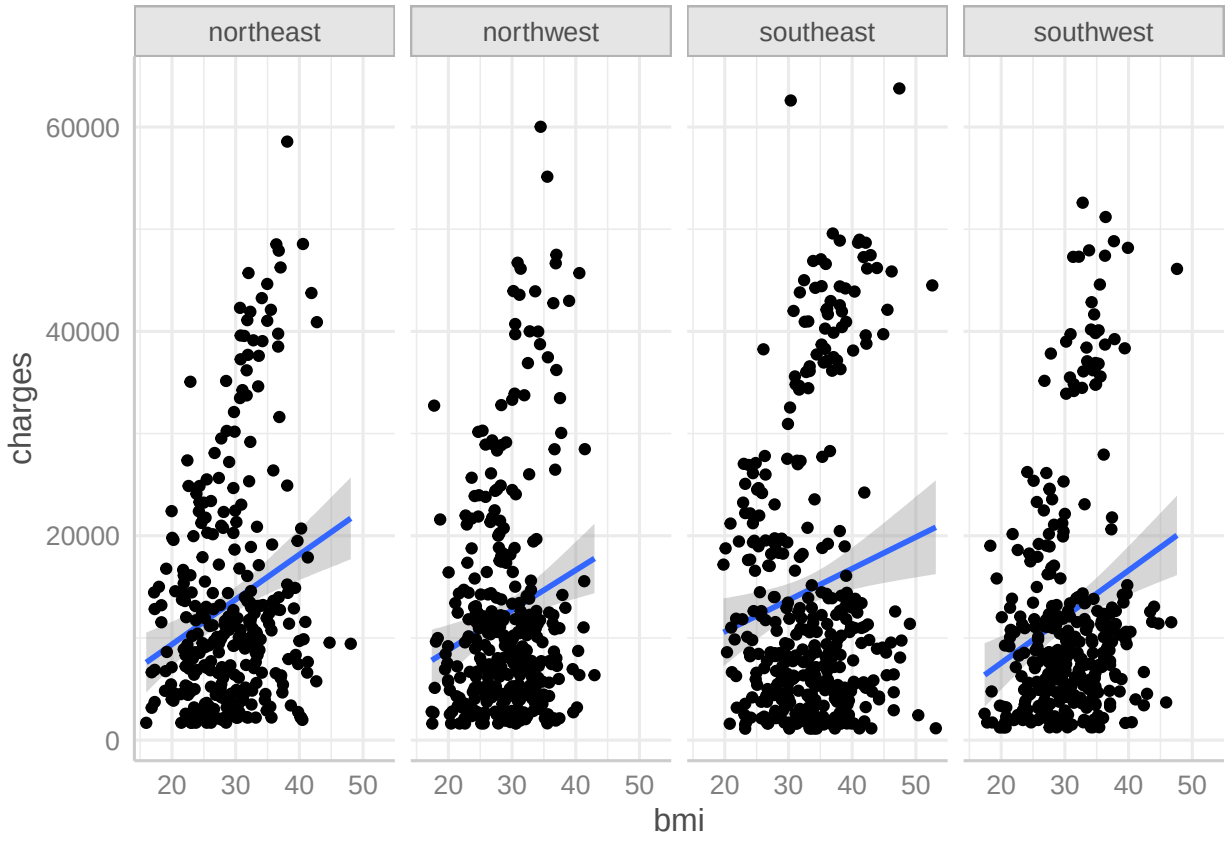
```
library(ggplot2)

# charges-age relationship, grouped by region
ggplot(Data,aes(x=age,y=charges)) + geom_smooth(method = "lm",level = 0.95) +
geom_point() + facet_wrap(~region, nrow = 1, ncol = 4)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



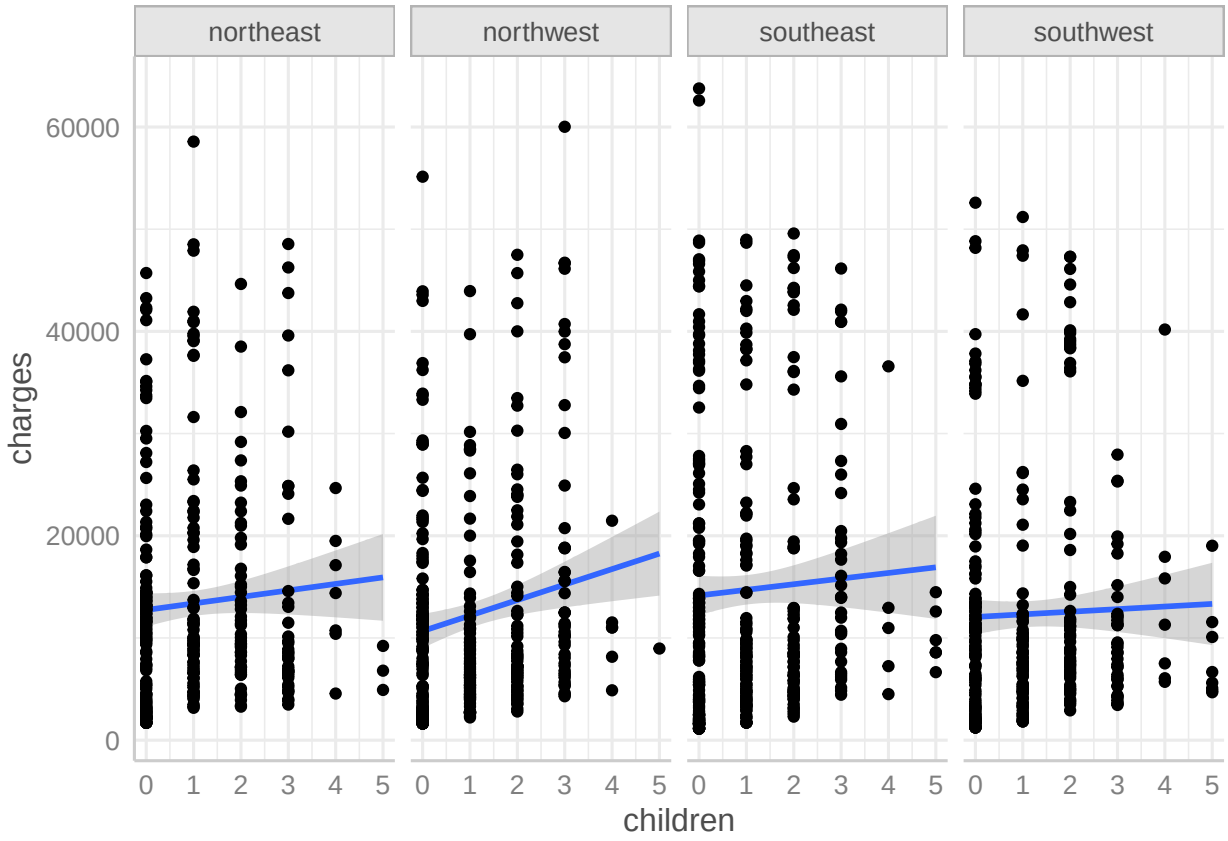

```
# charges-bmi relationship, grouped by region
ggplot(Data,aes(x=bmi,y=charges)) + geom_smooth(method = "lm",level = 0.95) +
geom_point() + facet_wrap(~region, nrow = 1, ncol = 4)
```

```
## `geom_smooth()` using formula 'y ~ x'
```



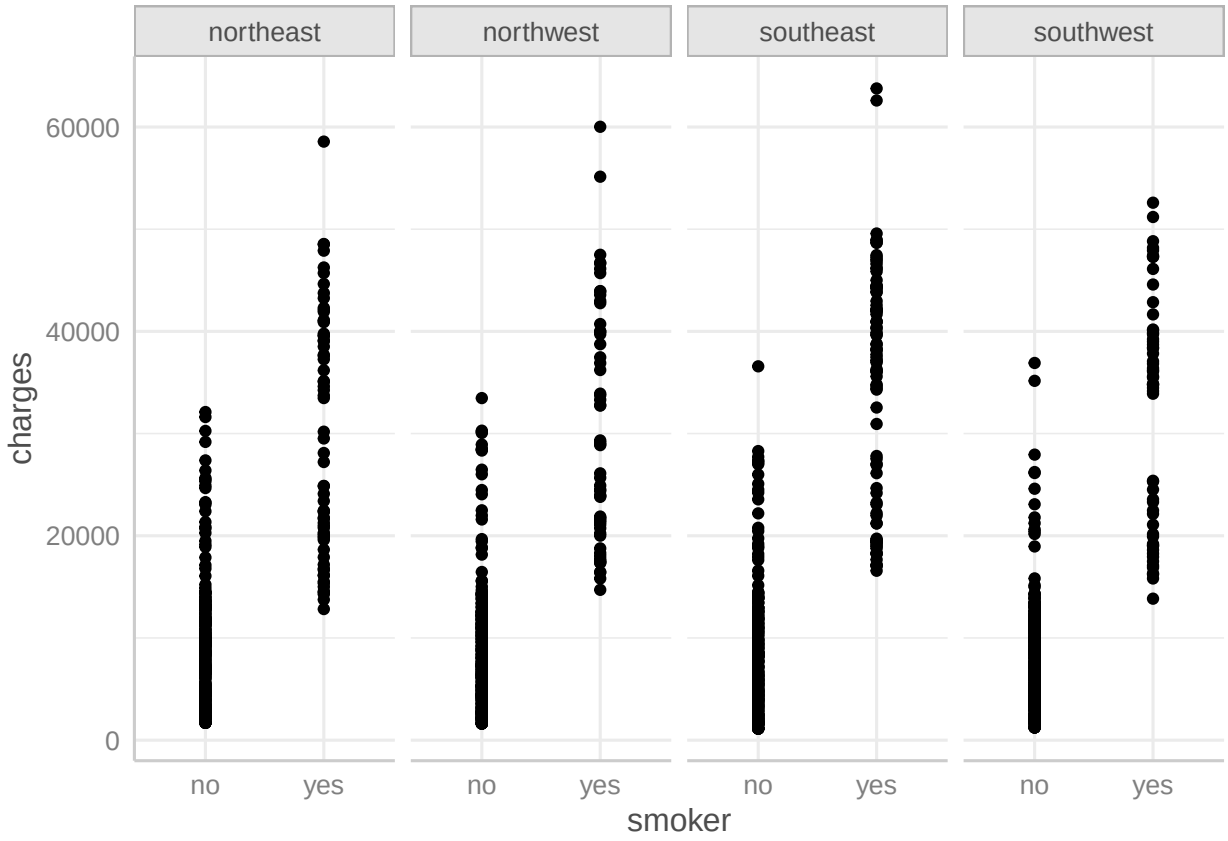

```
# charges-children relationship, grouped by region
ggplot(Data,aes(x=children,y=charges)) + geom_smooth(method = "lm",level = 0.95) +
geom_point() + facet_wrap(~region, nrow = 1, ncol = 4)
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
# charges-smoker relationship, grouped by region
ggplot(Data,aes(x=smoker,y=charges)) + geom_smooth(method = "lm",level = 0.95) +
geom_point() + facet_wrap(~region, nrow = 1, ncol = 4)
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
library(lme4)
```

```
## Loading required package: Matrix
```

```
## Warning: package 'Matrix' was built under R version 4.1.1
```

```
##
## Attaching package: 'Matrix'
```

```
## The following objects are masked from 'package:tidyr':
##
##     expand, pack, unpack
```

```
model_lr <- lm(charges~sex+age+bmi,data=insurance)
model_lmer <- lmer(charges~sex+age+bmi+(1|region), data=insurance)
anova(model_lmer, model_lr)
```

```
## refitting model(s) with ML (instead of REML)
```

```
## Data: insurance
## Models:
## model_lr: charges ~ sex + age + bmi
## model_lmer: charges ~ sex + age + bmi + (1 | region)
##            npar   AIC   BIC logLik deviance  Chisq Df Pr(>Chisq)
## model_lr      5 28794 28820 -14392    28784
## model_lmer    6 28796 28827 -14392    28784 0.0541  1      0.816
```

```
summary(model_lmer)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: charges ~ sex + age + bmi + (1 | region)
##    Data: insurance
##
## REML criterion at convergence: 28737.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.3301 -0.6210 -0.4400  0.5972  4.1444
##
## Random effects:
##  Groups   Name        Variance  Std.Dev.
##  region   (Intercept)    232625   482.3
##  Residual             129135439 11363.8
## Number of obs: 1338, groups:  region, 4
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) -6873.71    1794.59  -3.830
## sexmale      1341.20     622.25   2.155
## age           243.57      22.27  10.938
## bmi           323.17      52.17   6.194
##
## Correlation of Fixed Effects:
##         (Intr) sexmal age
## sexmale -0.145
## age     -0.390  0.026
## bmi     -0.827 -0.048 -0.113
```

```
ranef(model_lmer)
```

```
## $region
##           (Intercept)
## northeast    225.7956
## northwest   -131.7809
## southeast    256.8350
## southwest   -350.8496
##
## with conditional variances for "region"
```

```
insurance$charges_cat <- ifelse(insurance$charges >median(insurance$charges), 1, 0)

insurance$smoker <- as.factor(insurance$smoker)

model_glmer<-glmer(charges_cat~sex+age+bmi+children+smoker+(1|region),
                   family=binomial,
                   data=insurance)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.00346307 (tol = 0.002, component 1)
```

```
## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, : Model is nearly unider
##  - Rescale variables?
```

```
summary(model_glmer)
```

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: charges_cat ~ sex + age + bmi + children + smoker + (1 | region)
##    Data: insurance
##
##      AIC      BIC   logLik deviance df.resid
##    796.4    832.8   -391.2    782.4     1331
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.4007 -0.2968 -0.0323  0.2958 13.6723
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  region (Intercept) 0.07901  0.2811
## Number of obs: 1338, groups:  region, 4
##
## Fixed effects:
##               Estimate Std. Error  z value Pr(>|z|)
## (Intercept) -8.558e+00  2.050e-02 -417.477  < 2e-16 ***
## sexmale     -3.477e-01  2.038e-02  -17.058  < 2e-16 ***
## age          1.663e-01  6.423e-03   25.883  < 2e-16 ***
## bmi          2.907e-02  9.381e-03    3.099  0.00194 **
## children     1.438e-01  1.972e-02    7.293 3.03e-13 ***
## smokeryes    2.194e+05  2.560e+02  857.102  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) sexmal age    bmi    chldrn
## sexmale    0.000
## age       -0.033 -0.019
## bmi       -0.031 -0.013 -0.843
## children  -0.003 -0.001 -0.006 -0.068
## smokeryes  0.000  0.000  0.000  0.000  0.000
## optimizer (Nelder_Mead) convergence code: 0 (OK)
## Model failed to converge with max|grad| = 0.00346307 (tol = 0.002, component 1)
## Model is nearly unidentifiable: large eigenvalue ratio
##  - Rescale variables?
```