Abs e3

```
library(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
```

```
library(ggplot2)
library(statsr)
library(grid) #Grid plotting plotting side by side or vertical
library(gridExtra) #neede for gri,arrange
```

```
##
## Attaching package: 'gridExtra'

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

```
library(tidyr)
library(scales)
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------------------

## v tibble  3.0.3     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
## v purrr   0.3.4

## -- Conflicts ------------------------------------------------------------------------------------
## x readr::col_factor()  masks scales::col_factor()
## x gridExtra::combine() masks dplyr::combine()
## x purrr::discard()     masks scales::discard()
## x dplyr::filter()      masks stats::filter()
## x dplyr::lag()         masks stats::lag()
```

```r
library(broom)
library(devtools)
```

```
## Loading required package: usethis
```

```r
library(lattice)
library(viridis)
```

```
## Loading required package: viridisLite
```

```
##
## Attaching package: 'viridis'
```

```
## The following object is masked from 'package:scales':
##
##     viridis_pal
```

```r
Abs = read.csv('Absentees Dataset.csv')
Ab = as.matrix(Abs)
str(Abs)
```

Reading data

```
## 'data.frame':    740 obs. of  22 variables:
##  $ ID                              : num  11 36 3 7 11 3 10 20 14 1 ...
##  $ Reason.for.absence              : num  26 0 23 7 23 23 22 23 19 22 ...
##  $ Month.of.absence                : num  7 7 7 7 7 7 7 7 7 7 ...
##  $ Day.of.the.week                 : num  3 3 4 5 5 6 6 6 2 2 ...
##  $ Seasons                         : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ Transportation.expense          : num  289 118 179 279 289 179 361 260 155 235 ...
##  $ Distance.from.Residence.to.Work: num  36 13 51 5 36 51 52 50 12 11 ...
##  $ Service.time                    : num  13 18 18 14 13 18 3 11 14 14 ...
##  $ Age                             : num  33 50 38 39 33 38 28 36 34 37 ...
##  $ Work.load.Average.day           : num  240 240 240 240 240 ...
##  $ Hit.target                      : num  97 97 97 97 97 97 97 97 97 97 ...
##  $ Disciplinary.failure            : num  0 1 0 0 0 0 0 0 0 0 ...
##  $ Education                       : num  1 1 1 1 1 1 1 1 1 3 ...
##  $ Son                             : num  2 1 0 2 2 0 1 4 2 1 ...
##  $ Social.drinker                  : num  1 1 1 1 1 1 1 1 1 0 ...
##  $ Social.smoker                   : num  0 0 0 1 0 0 0 0 0 0 ...
##  $ Pet                             : num  1 0 0 0 1 0 4 0 0 1 ...
##  $ Weight                          : num  90 98 89 68 90 89 80 65 95 88 ...
##  $ Height                          : num  172 178 170 168 172 170 172 168 196 172 ...
##  $ Body.mass.index                 : num  30 31 31 24 30 31 27 23 25 29 ...
##  $ Absenteeism.time.in.hours       : num  4 0 2 4 2 2 8 4 40 8 ...
##  $ icd_attested                    : int  0 1 0 1 0 0 0 0 1 0 ...
```

```
Abs <- Abs%>%
  group_by(ID)%>%
  mutate(Num_Abs = n())

#note it was Abs__single single

Abs <- Abs%>%
  group_by(ID)%>%
  mutate(Sum_Abs = sum(Absenteeism.time.in.hours))

Abs_single <-Abs[!duplicated(Abs$ID), ] #removing repeated ID

#Abs_single$ID<- as.factor(Abs_single$ID)

Abs_single$ID <- factor(Abs_single$ID,
                  levels = Abs_single$ID[order(Abs_single$Sum_Abs, decreasing = FALSE)])
```

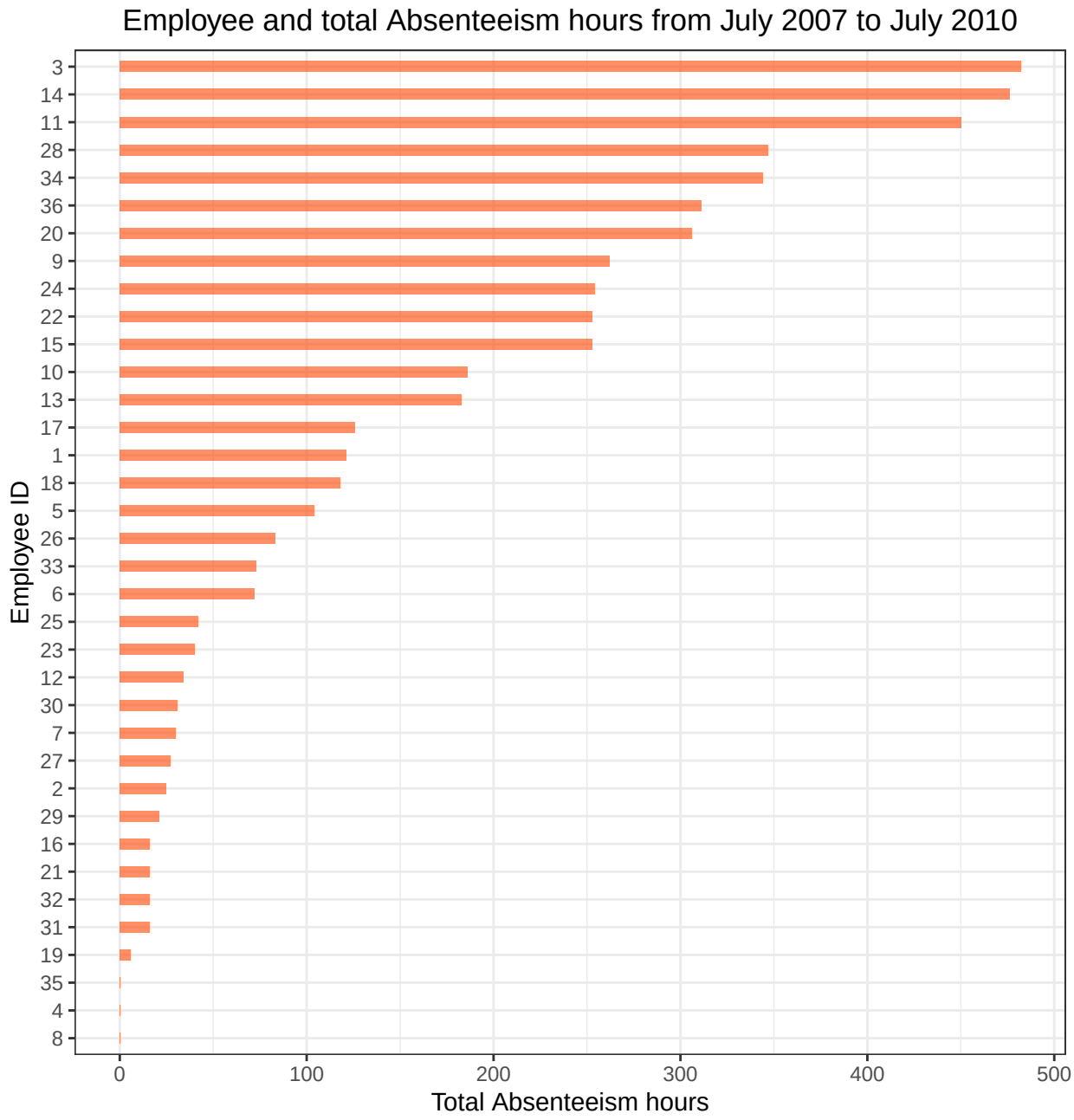
labs(title = “Reason for absence (ICD)”,

```
   y = "Frequency" ,x ="Life style") +
```

theme_bw()

```
theme_bw()
```

```
Abs_single %>%


  ggplot( aes(x=ID, y=Sum_Abs)) +
    geom_bar(stat="identity", fill="orangered1", alpha=.6, width=.4) +
    coord_flip() +

  ylab("Total Absenteeism hours")+
  xlab("Employee ID")+
ggtitle("Employee and total Absenteeism hours from July 2007 to July 2010")+
  labs(caption = "3 is the most common reason for absence which is Diseases of the blood and blood-form
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



eases of the blood and blood−forming organs and certain disorders involving the immune mechanism

```
#Abs_single <-Abs[!duplicated(Abs$ID), ] #removing repeated ID
```

Abs_single$ID <- -factor(Abs_single$ID,
levels = Abs_single$ID[order(Abs_single$Sum_Abs, decreasing = FALSE)])

Abs <- Abs%>% group_by(ID)%>% mutate(Num_Abs = n())

https://www.cdc.gov/healthyweight/assessing/index.html

#note single Abs_single <- Abs%>% group_by(ID)%>% mutate(Sum_Abs = sum(Absenteeism.time.in.hours))

```
Abs_single <- Abs_single %>%
  mutate(Obesity = ifelse(Body.mass.index < 18.5,"Underweight", ifelse(Body.mass.index < 24.9 ,"Normal
```

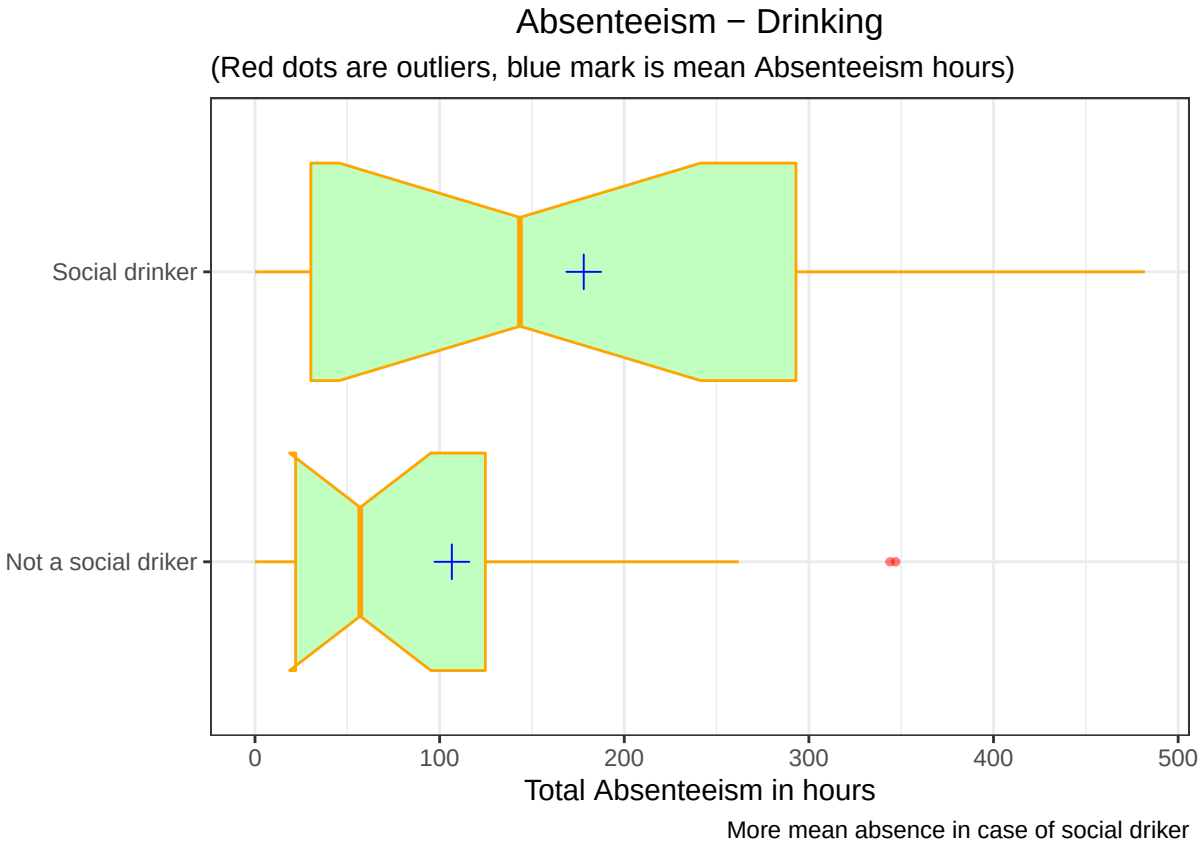
```
Abs_single %>%
  group_by(Obesity)%>%
  summarise(mean(Sum_Abs))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 3 x 2
##   Obesity       `mean(Sum_Abs)`
##   <chr>                   <dbl>
## 1 Normal Weight            125.
## 2 Obese                    162.
## 3 Overweight               147.
```

```
ggplot(data=Abs_single,

       aes(y= factor(Social.drinker,labels = c("Not a social driker","Social drinker"))

           ,x=Sum_Abs,color =Social.drinker ))+
  geom_boxplot(notch = TRUE,fill = "darkseagreen1", colour = "orange1",outlier.color = "red",
               outlier.alpha = 0.5,
               outlier.size = 1)+ stat_summary(fun=mean, geom="point", shape=3, size=4, color="blue")+


  scale_x_continuous(labels = scales::comma)+

  labs(title="Absenteeism - Drinking",subtitle = "(Red dots are outliers, blue mark is mean Absenteeism
       x = "Total Absenteeism in hours",y="")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```

```
## notch went outside hinges. Try setting notch=FALSE.
```

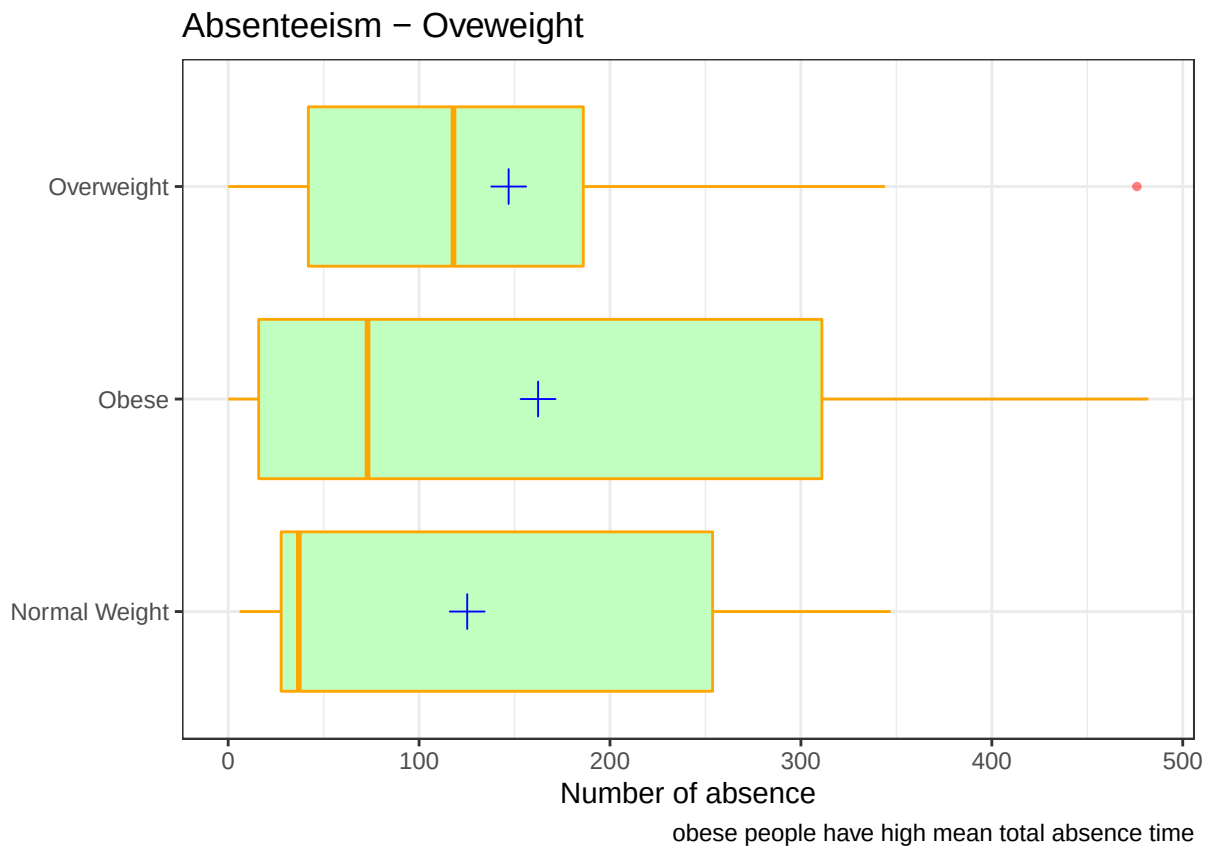



```
Abs_single %>%
  group_by(Obesity)%>%
  summarise(mean(Sum_Abs))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 3 x 2
##   Obesity       `mean(Sum_Abs)`
##   <chr>                   <dbl>
## 1 Normal Weight            125.
## 2 Obese                    162.
## 3 Overweight               147.
```

```
ggplot(data=Abs_single,

       aes(y= factor(Obesity,labels = c("Normal Weight","Obese","Overweight"))

           ,x=Sum_Abs,color =Obesity ))+
  geom_boxplot(fill = "darkseagreen1", colour = "orange",outlier.color = "red",
               outlier.alpha = 0.5,
               outlier.size = 1)+
  stat_summary(fun=mean, geom="point", shape=3, size=4, color="blue")+


  scale_x_continuous(labels = scales::comma)+
```

```
  labs(title="Absenteeism - Oveweight", caption = "obese people have high mean total absence time",
       x = "Number of absence",y="")+
  theme_bw()
```



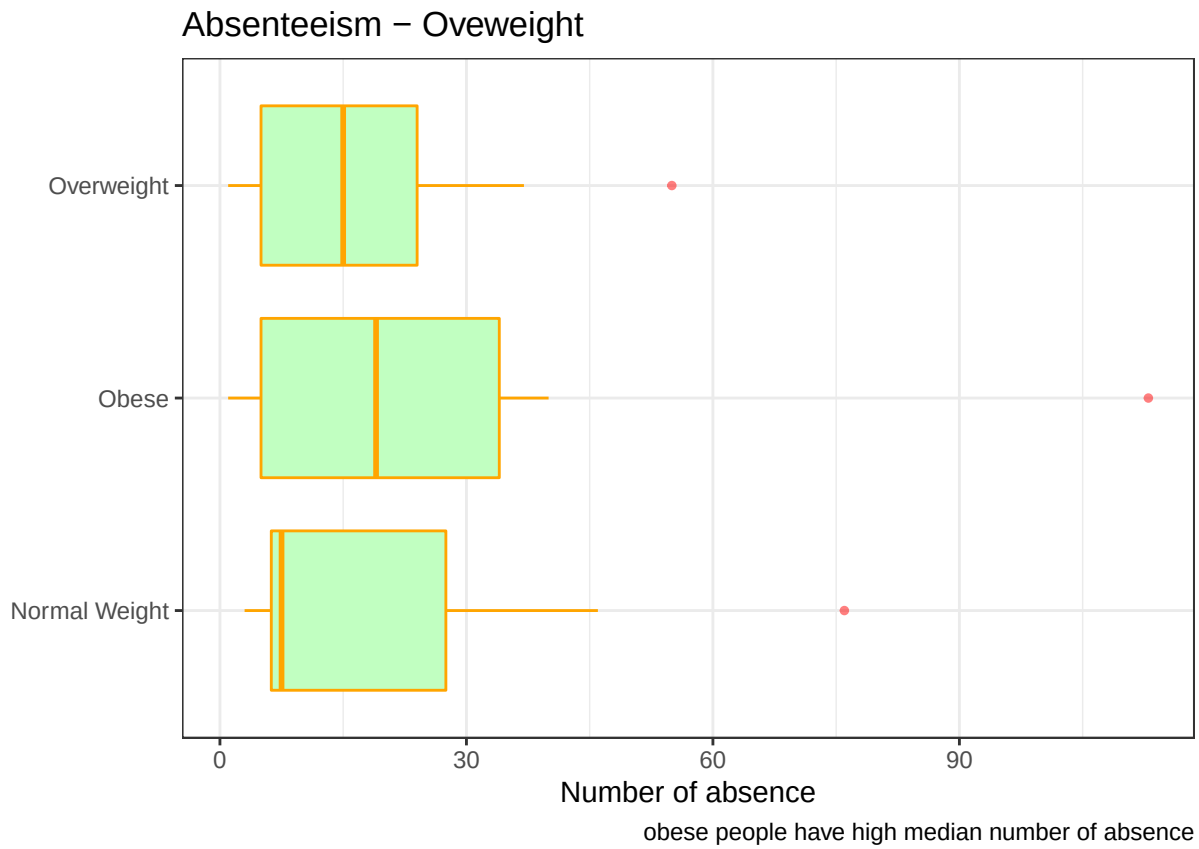

```
Abs_single %>%
  group_by(Obesity)%>%
  summarise(median(Num_Abs))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 3 x 2
##   Obesity       `median(Num_Abs)`
##   <chr>                     <dbl>
## 1 Normal Weight               7.5
## 2 Obese                      19
## 3 Overweight                 15
```

```
ggplot(data=Abs_single,

       aes(y= factor(Obesity,labels = c("Normal Weight","Obese","Overweight"))

           ,x=Num_Abs,color =Obesity ))+
  geom_boxplot(fill = "darkseagreen1", colour = "orange",outlier.color = "red",
```

```
              outlier.alpha = 0.5,
              outlier.size = 1)+


  scale_x_continuous(labels = scales::comma)+

  labs(title="Absenteeism - Oveweight", caption = "obese people have high median number of absence",
       x = "Number of absence",y="")+
  theme_bw()
```




```
Abs_single <- Abs_single %>%
  mutate(Service_grp = ifelse(Service.time < 5,"Less than 5 years of service",ifelse(Service.time < 15,
```

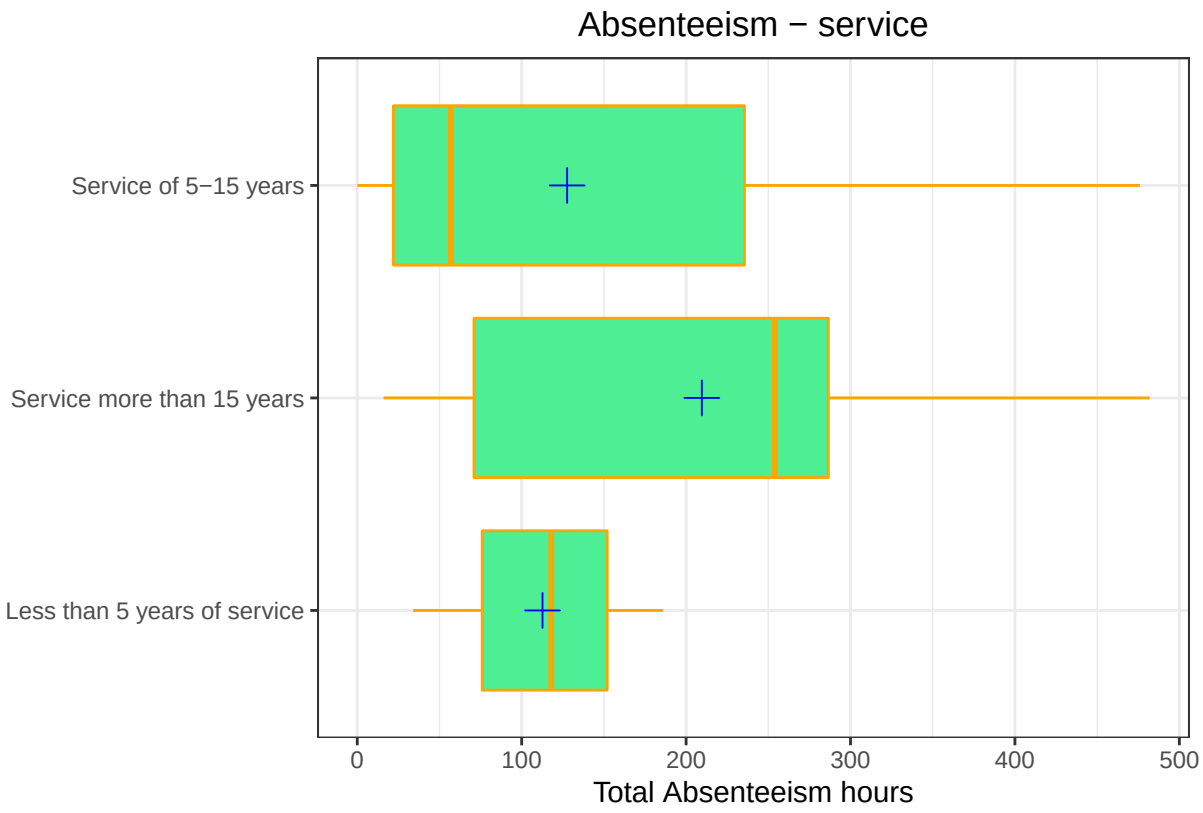
Abs_single <- Abs_single %>% group_by(Service_grp)%>% mutate(grp = mean(Sum_Abs))

```
Abs_single %>%
  group_by(Service_grp)%>%
  summarise(mean(Sum_Abs))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 3 x 2
##   Service_grp                  `mean(Sum_Abs)`
##   <chr>                                  <dbl>
```

```
## 1 Less than 5 years of service            113.
## 2 Service more than 15 years              210.
## 3 Service of 5-15 years                   128.
```

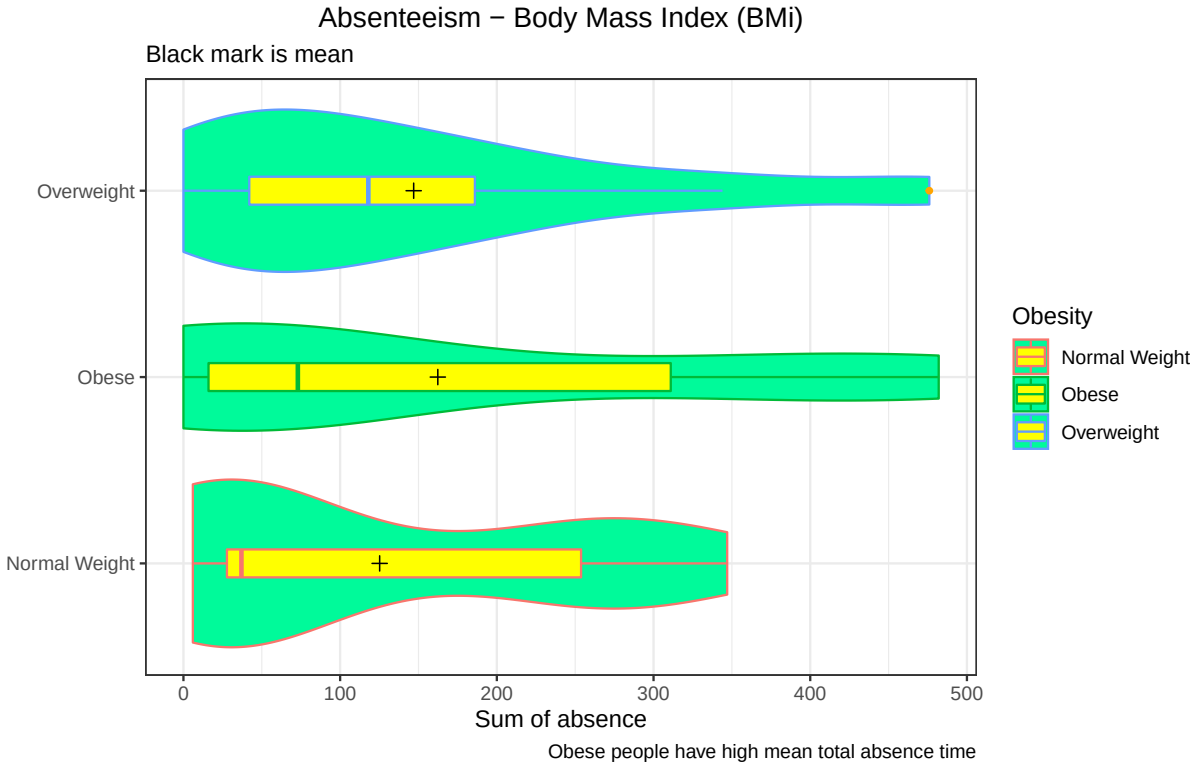
```
ggplot(data=Abs_single,

       aes(y= Service_grp

           ,x=Sum_Abs,color =Service_grp ))+
  geom_boxplot(fill = "seagreen2", colour = "orange",outlier.color = "red",
               outlier.alpha = 0.5,
               outlier.size = 1)+stat_summary(fun=mean, geom="point", shape=3, size=4, color="blue")+


  scale_x_continuous(labels = scales::comma)+

  labs(title="Absenteeism - service", caption = "People with more than 15 years of service have high mea
       x = "Total Absenteeism hours",y="")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



People with more than 15 years of service have high mean Total Absenteeism hours

```
ggplot(data=Abs_single,

       aes(y= factor(Obesity,labels = c("Normal Weight","Obese","Overweight"))
```

```
          ,x=Sum_Abs,color =Obesity ))+
  geom_violin(fill = "mediumspringgreen") +
  geom_boxplot(width = .15,
               fill = "yellow",
               outlier.color = "orange",
               outlier.size = 1) + stat_summary(fun=mean, geom="point", shape=3, size=2, color="black")+
  labs(title="Absenteeism - Body Mass Index (BMi)",subtitle ="Black mark is mean", caption = "Obese peop
       y = "",x="Sum of absence")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```




```
Abs %>%
  group_by(Month.of.absence)%>%
  mutate(Abs_month= Absenteeism.time.in.hours>0)%>%
  summarize(Abs_month)
```

```
## `summarise()` regrouping output by 'Month.of.absence' (override with `.groups` argument)

## # A tibble: 740 x 2
## # Groups:   Month.of.absence [13]
##    Month.of.absence Abs_month
##               <dbl> <lgl>
##  1                0 FALSE
##  2                0 FALSE
##  3                0 FALSE
##  4                1 TRUE
##  5                1 TRUE
```

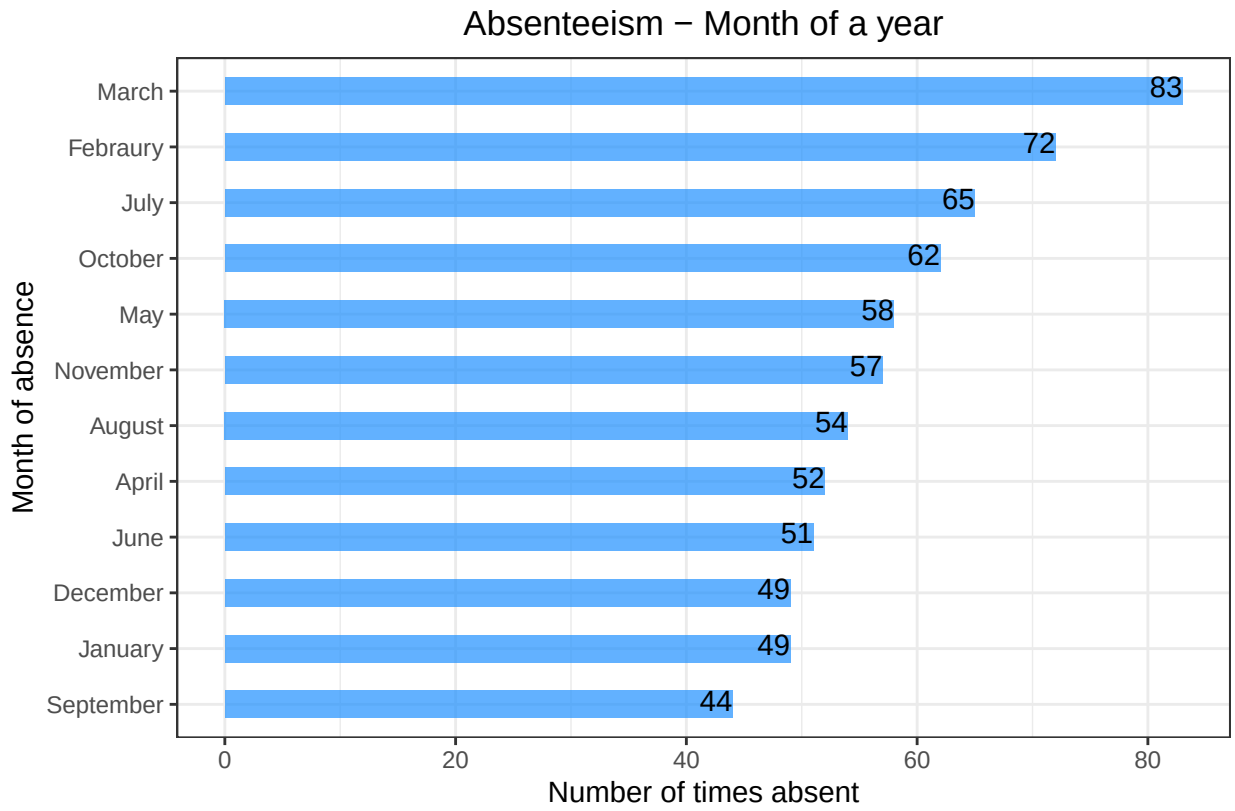
```
##  6                1 TRUE
##  7                1 TRUE
##  8                1 TRUE
##  9                1 TRUE
## 10                1 TRUE
## # ... with 730 more rows
```

```
library(scales)
A <-Abs %>%
  group_by(Month.of.absence)%>%
  filter(Absenteeism.time.in.hours>0)%>%
  summarise(n=n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
A<- A%>%
  mutate(Month= ifelse(Month.of.absence=="1","January",ifelse(Month.of.absence=="2","Febraury",ifelse(M
```

```
A$Month <-  factor(A$Month,
                   levels = A$Month[order(A$n, decreasing = FALSE)])

A%>%

ggplot(aes(x = Month, y = n) )+
  geom_bar( stat="identity", fill="dodgerblue1", alpha=.7, width=.5)+
  geom_text(aes(label = n),
            vjust = +0.30,hjust=1.0) +
    coord_flip() +
  labs(title="Absenteeism - Month of a year", caption = "High number of absence in the month of march",
       y = "Number of times absent",x="Month of absence")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



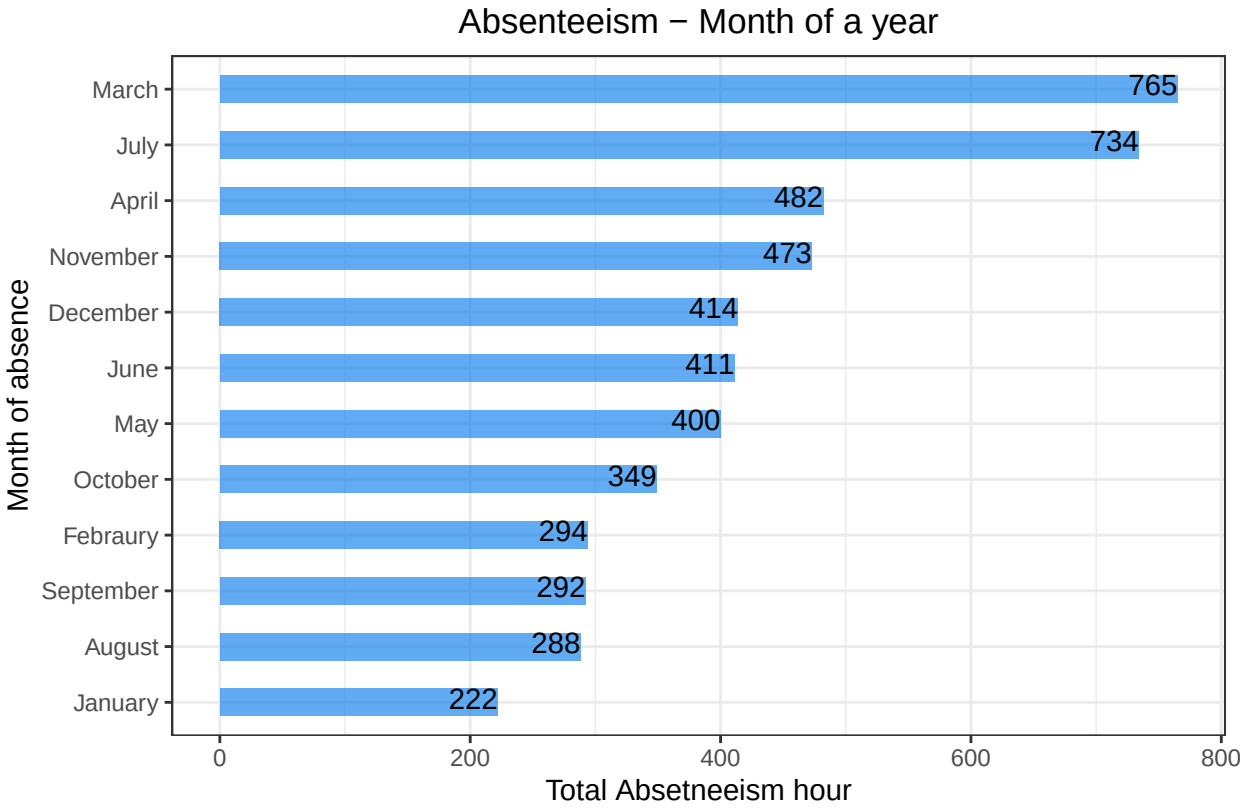
High number of absence in the month of march

```
B <-Abs %>%
  group_by(Month.of.absence)%>%
  filter(Absenteeism.time.in.hours>0)%>%
  summarize(Sum_abs_month= sum(Absenteeism.time.in.hours))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
 B<- B%>%
  mutate(Month= ifelse(Month.of.absence=="1","January",ifelse(Month.of.absence=="2","Febraury",ifelse(M
```

```
B$Month <-  factor(B$Month,
                   levels = B$Month[order(B$Sum_abs_month, decreasing = FALSE)])

B%>%

ggplot(aes(x = Month, y = Sum_abs_month) )+
  geom_bar( stat="identity", fill="dodgerblue2", alpha=.7, width=.5)+
  geom_text(aes(label = Sum_abs_month),
            vjust = +0.30,hjust=1.0) +
    coord_flip()+  labs(title="Absenteeism - Month of a year", caption = "Large absenteeism hours in th
       y = "Total Absetneeism hour",x="Month of absence")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



Large absenteeism hours in the month of march

```
q<-Abs_single%>%
  group_by(Distance.from.Residence.to.Work)%>%
  summarise(Sum_Abs)
```

```
## `summarise()` regrouping output by 'Distance.from.Residence.to.Work' (override with `.groups` argumer
```

```
w <- Abs_single %>%
  mutate(Distance_grp = ifelse(Distance.from.Residence.to.Work < 15,"Less than 15kms Distance.from.Resi
```

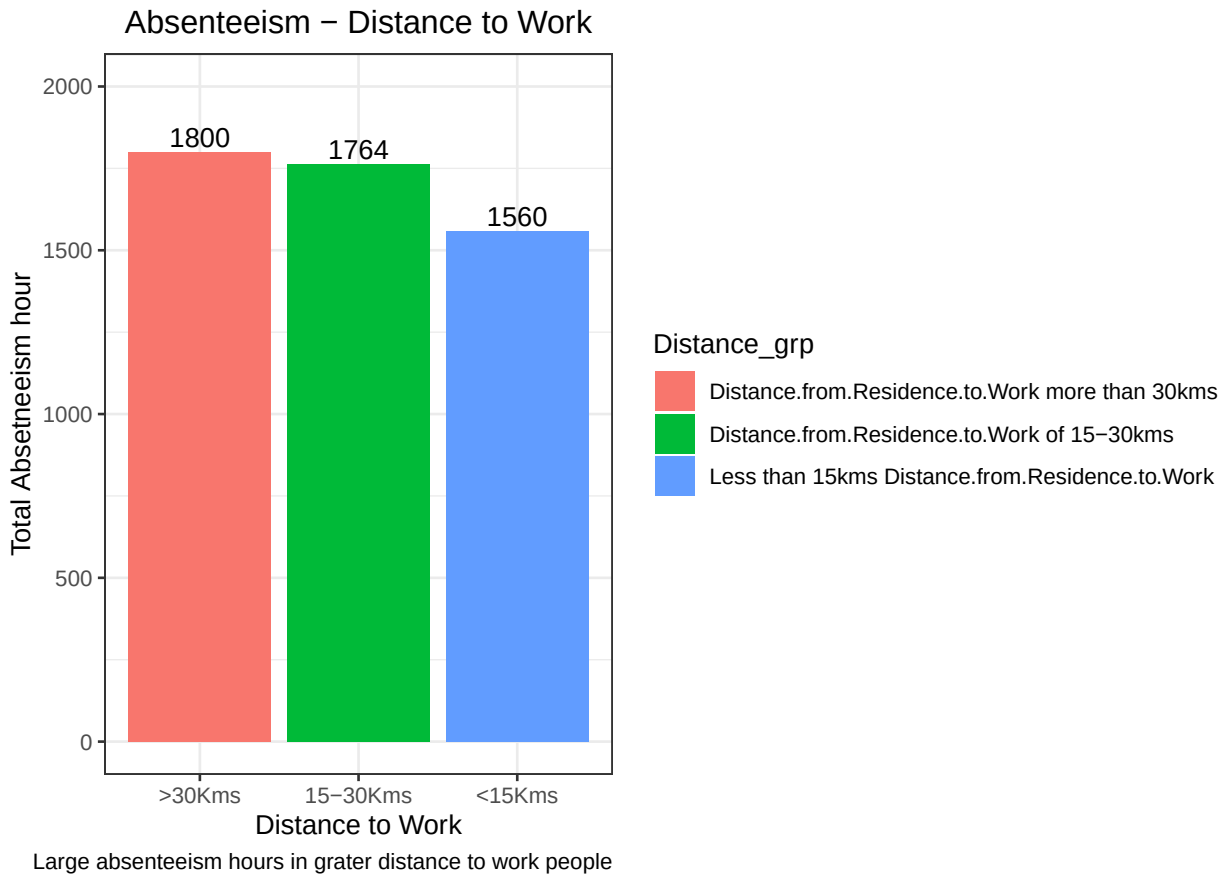
```
e<-w%>%
  group_by(Distance_grp)%>%
  summarise(total_Abs_time_distance = sum(Sum_Abs))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

- labs(x = "Distance to work", y = "Total absence hour", title = "Absenteeism and distance to work")

```
e%>%
ggplot(aes(x = factor(Distance_grp,label=c(">30Kms","15-30Kms","<15Kms")),y=total_Abs_time_distance,fil
  geom_bar(stat = "identity")+
   geom_text(aes(label = total_Abs_time_distance),
             vjust = -0.25)+
  scale_y_continuous(limits=c(0,2000))+
```

```
  labs(title="Absenteeism - Distance to Work", caption = "Large absenteeism hours in grater distance to
       y = "Total Absetneeism hour",x="Distance to Work")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



Large absenteeism hours in grater distance to work people

factor(Distance_grp,labels = c("Distance.from.Residence.to.Work of 15-30kms","Less than 15kms Distance.from.Residence.to.Work","Distance.from.Residence.to.Work more than 30kms")

No grouping of ID is done from now on

```
plotdata_2 <- Abs %>%
  group_by(Day.of.the.week) %>%
  summarize(Average_abs = mean(Absenteeism.time.in.hours))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

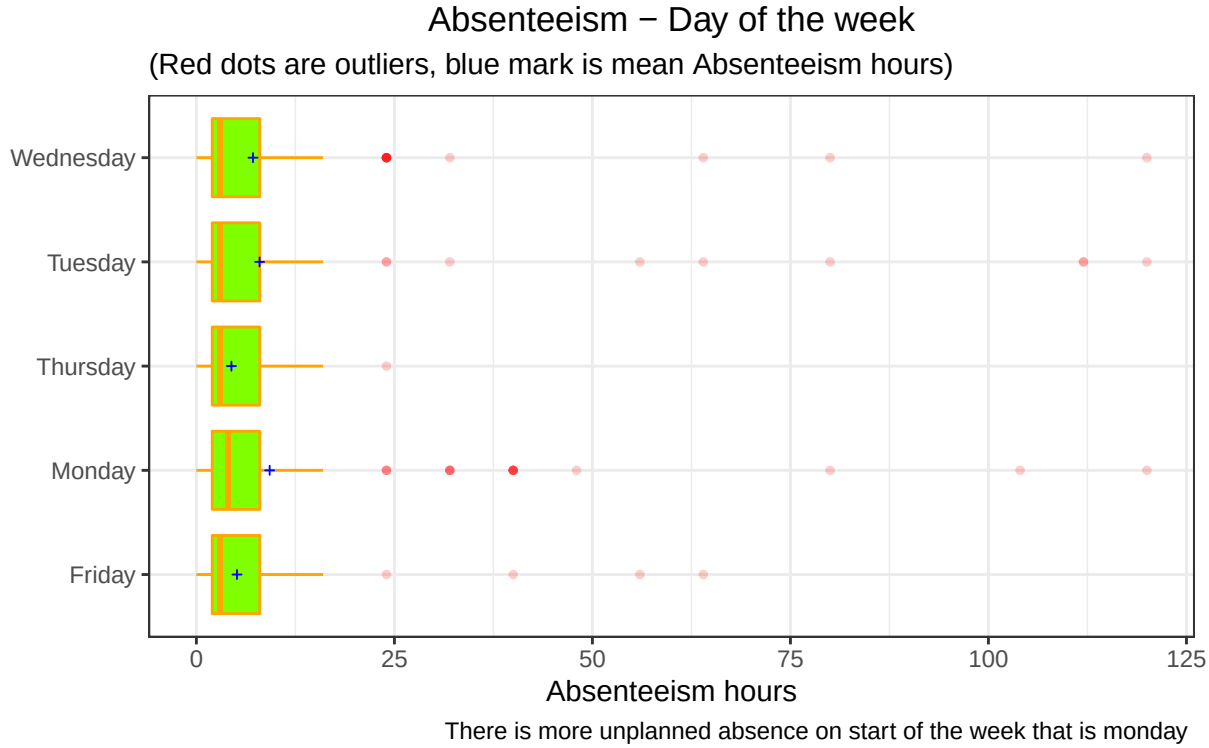
```
p1<-ggplot(plotdata_2,
       aes(x = factor(Day.of.the.week),
                      y = Average_abs)) +
  geom_bar(stat = "identity",
           fill = "chartreuse1" ) +
 geom_text( aes(label = Average_abs),
```

```
            vjust = -0.25)+

  scale_y_continuous(breaks = seq(0, 20, 1)) +
  labs(title = "Day of the week average Absenteeism time in hours ",
       subtitle = "Absenteeism_at_work data-set Cascade cup",
       x = "",
       y = "Mean absenteeism time in hours")
```

```
Abs<-Abs%>%
  mutate(Day = ifelse(Day.of.the.week==2,"Monday",ifelse(Day.of.the.week==3,"Tuesday",ifelse(Day.of.the
```

,labels = c("Monday","Tuesday","Wednesday","Thursday", "Friday"

```
ggplot(data=Abs,

       aes(y= Day

           ,x=Absenteeism.time.in.hours,color =Day ))+
  geom_boxplot(fill = "chartreuse1", colour = "orange",outlier.color = "red",
               outlier.alpha = 0.2,
               outlier.size = 1)+ stat_summary(fun=mean, geom="point", shape=3, size=0.8, color="blue")+

  labs(title="Absenteeism - Day of the week",subtitle = "(Red dots are outliers, blue mark is mean Absen
       x = "Absenteeism hours",y="")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```

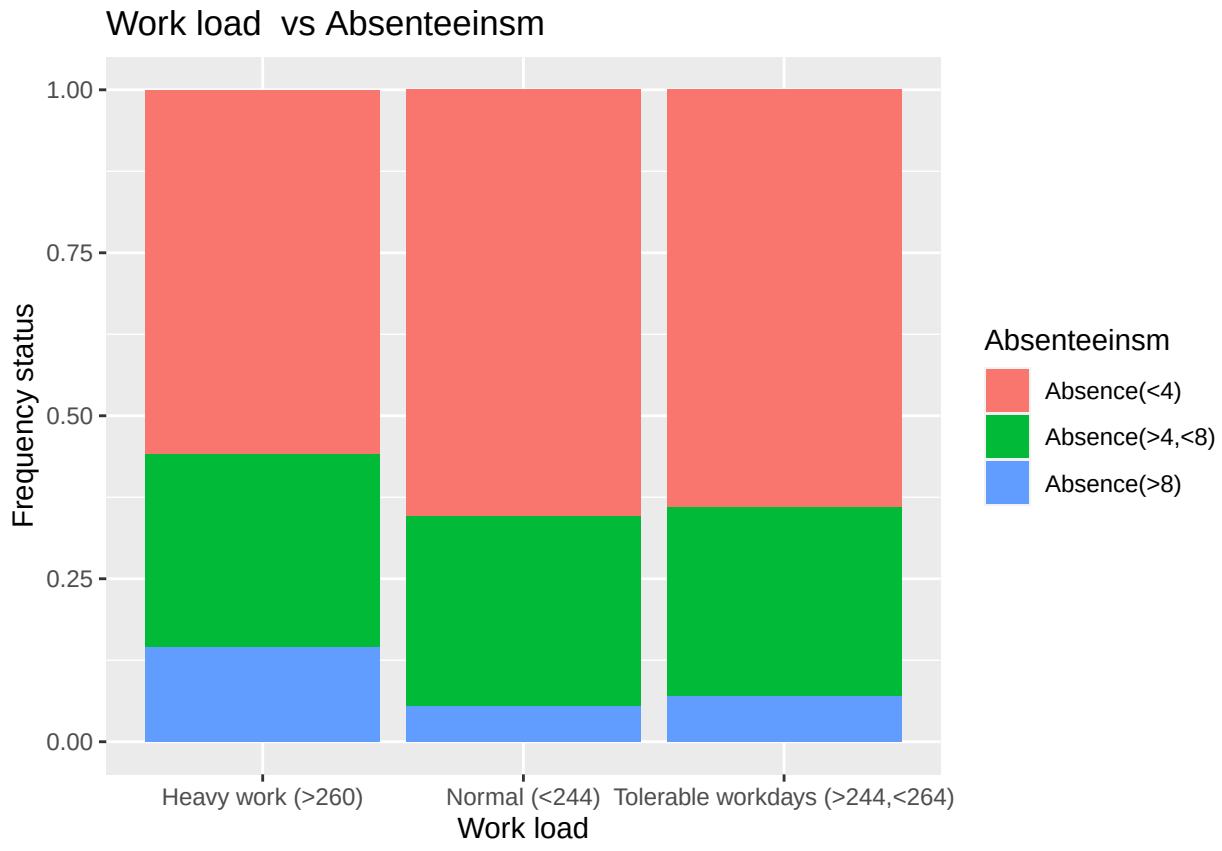


There is more unplanned absence on start of the week that is monday

```
#grid.arrange(p1, p2, ncol = 1)
```

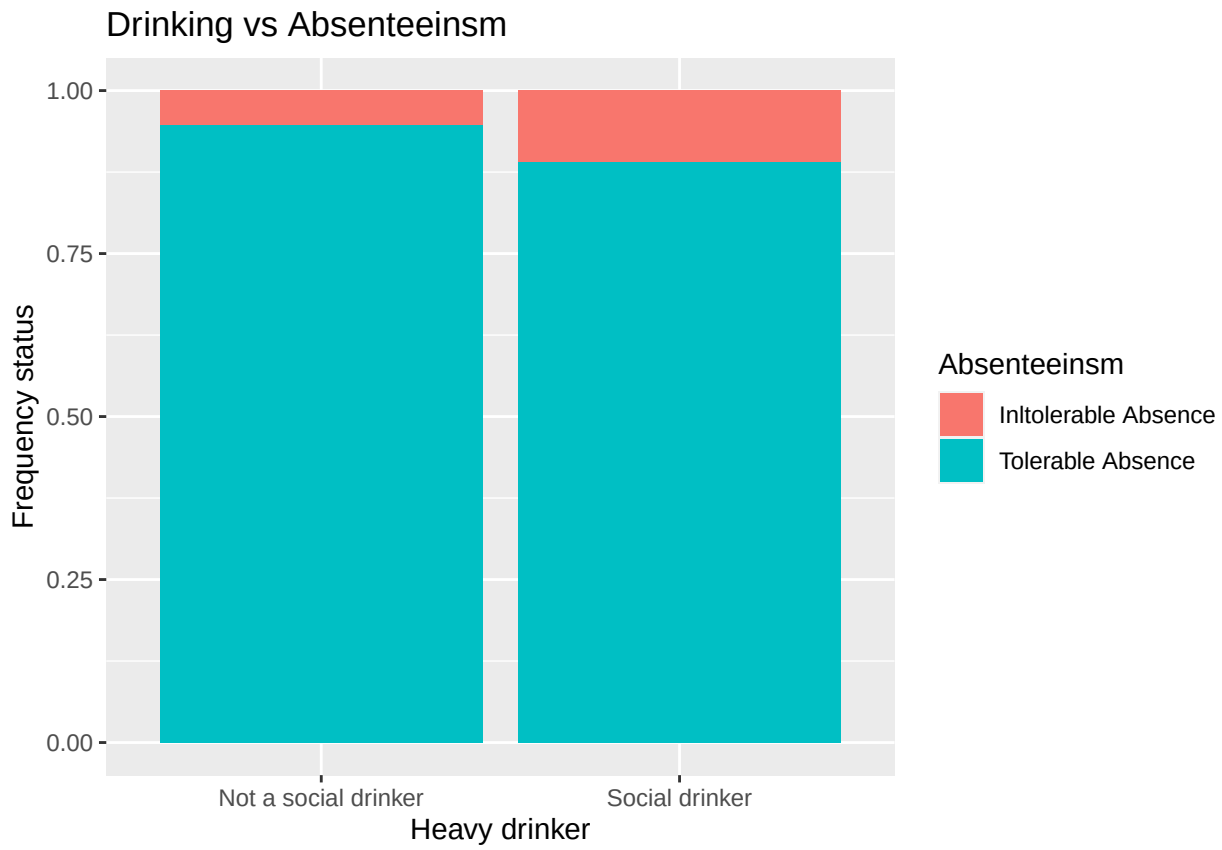
I dont think below one is fully accurate bcz, Why is Work load Average/day is so less? (Assuming that its given in minutes...max is 378=>6+hours) Or am I misunderstanding this ??

```
Abs <- Abs %>%
  mutate(Tol_Intol_above = ifelse(Absenteeism.time.in.hours > 8,"Absence(>8) ",ifelse(Absenteeism.time.

Abs <- Abs %>%
  mutate(Heavy_normal = ifelse(Work.load.Average.day > 294,"Heavy work (>260)",ifelse(Work.load.Average

ggplot(data = Abs, aes(x =Heavy_normal, fill = Tol_Intol_above)) +
  geom_bar(position="fill")+
 labs(
        x = "Work load",
        y = "Frequency status", fill = "Absenteeinsm" ,
        title = paste(
            "Work load  vs Absenteeinsm"
      ))
```




```
Abs <- Abs %>%
  mutate(Tol_Intol = ifelse(Absenteeism.time.in.hours > 8,"Inltolerable Absence", "Tolerable Absence"))

ggplot(data = Abs, aes(x =factor(Social.drinker,labels = c("Not a social drinker",
```

```
                                  "Social drinker")), fill = Tol_Intol)) +
 geom_bar(position="fill")+
labs(
       x = "Heavy drinker",
       y = "Frequency status", fill = "Absenteeinsm" ,
       title = paste(
           "Drinking vs Absenteeinsm"
     ))
```

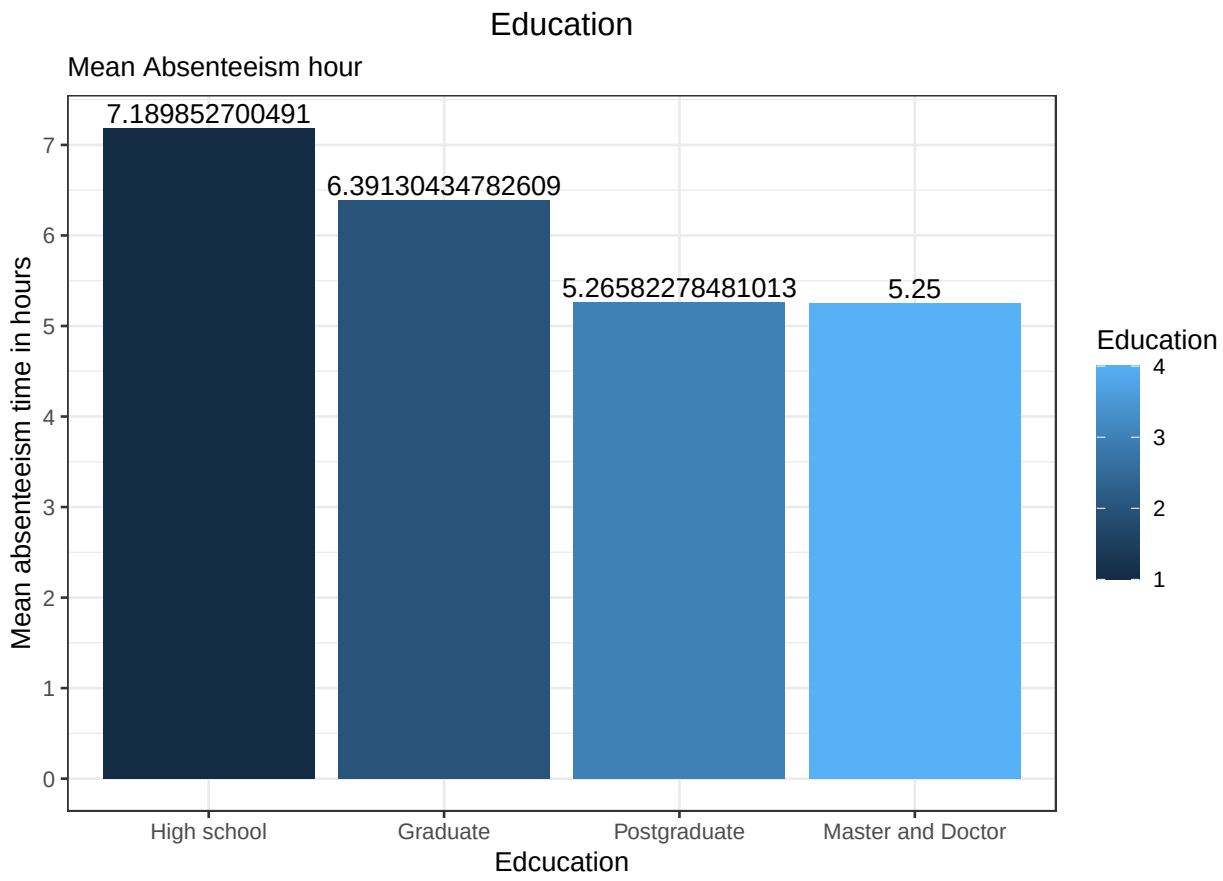



```
plotdata_3<-Abs%>%
  group_by(Education)%>%
  summarise(mean_abs_edu= mean(Absenteeism.time.in.hours))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
ggplot(plotdata_3,
       aes(x = factor(Education,label=c("High school", "Graduate", "Postgraduate", "Master and Doctor")),
                      y = mean_abs_edu,fill=Education)) +
  geom_bar(stat = "identity" ) +
 geom_text( aes(label = mean_abs_edu),
            vjust = -0.25)+

  scale_y_continuous(breaks = seq(0, 10, 1)) +
  labs(title = "Education",
```

```
      subtitle = "Mean Absenteeism hour",
      x = "Edcucation",
      y = "Mean absenteeism time in hours")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```




```
Abs%>%
  group_by(Social.drinker=="1" &Social.smoker=="1"&Tol_Intol=="Inltolerable Absence")%>%
  summarise(n=n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 2 x 2
##   `Social.drinker == "1" & Social.smoker == "1" & Tol_Intol == "Inltolera~     n
##   <lgl>                                                                    <int>
## 1 FALSE                                                                      737
## 2 TRUE                                                                         3
```

```
Abs<-Abs%>%
  mutate(Social.drinker_1=ifelse(Social.drinker=="1","Drinker" ,"Not a drinker"))
```

```
Abs<-Abs%>%
  mutate(Social.smoker_1= ifelse(Social.smoker=="1","Smoker","Not a smoker"))
```

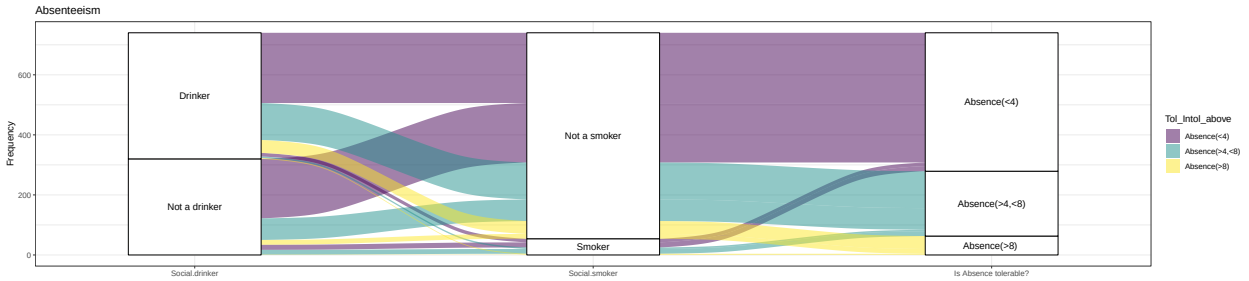
install.packages("ggalluvial")

```
library(ggalluvial)
```

```
## Warning: package 'ggalluvial' was built under R version 4.0.3
```

```
titanic_table <- Abs %>%
  group_by(Social.drinker_1, Social.smoker_1, Tol_Intol_above) %>%
  count()
```

```
ggplot(titanic_table,
       aes(axis1 = Social.drinker_1,
           axis2 = Social.smoker_1,
           axis3 = Tol_Intol_above,
           y = n)) +
  geom_alluvium(aes(fill=Tol_Intol_above)) +
  geom_stratum() +
  geom_text(stat = "stratum", aes(label = after_stat(stratum))) +
  scale_x_discrete(limits = c("Social.drinker", "Social.smoker", "Is Absence tolerable?"),
                   expand = c(.1, .2)) +
  scale_fill_viridis_d() +
  labs(title = "Absenteeism",

       y = "Frequency") +
  theme_bw()
```

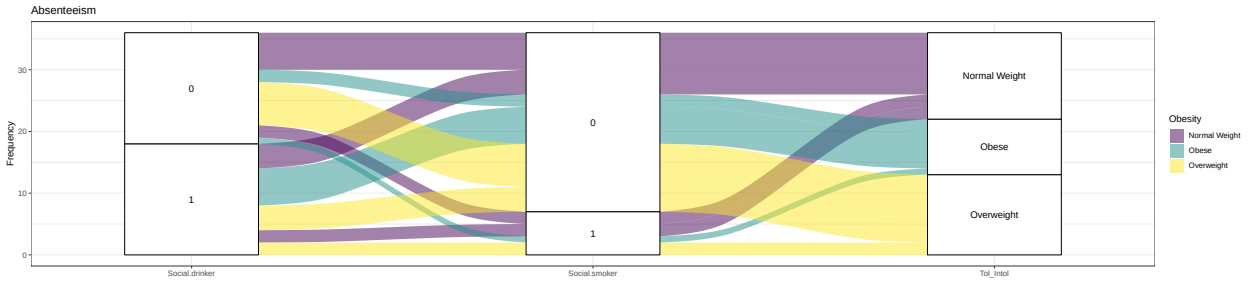



```
table_single<- Abs_single %>%
  group_by(Social.drinker, Social.smoker, Obesity) %>%
  count()
```

```
ggplot(table_single,
       aes(axis1 = Social.drinker,
           axis2 = Social.smoker,
           axis3 = Obesity,
           y = n)) +
  geom_alluvium(aes(fill=Obesity)) +
  geom_stratum() +
  geom_text(stat = "stratum", aes(label = after_stat(stratum))) +
```

```
  scale_x_discrete(limits = c("Social.drinker", "Social.smoker", "Tol_Intol"),
                   expand = c(.1, .2)) +
  scale_fill_viridis_d() +
  labs(title = "Absenteeism",

       y = "Frequency") +
  theme_bw()
```

```
## Warning in to_lodes_form(data = data, axes = axis_ind, discern =
## params$discern): Some strata appear at multiple axes.

## Warning in to_lodes_form(data = data, axes = axis_ind, discern =
## params$discern): Some strata appear at multiple axes.

## Warning in to_lodes_form(data = data, axes = axis_ind, discern =
## params$discern): Some strata appear at multiple axes.
```



Reason.for.absence

```
Abs$Reason.for.absence <-factor(Abs$Reason.for.absence)
```

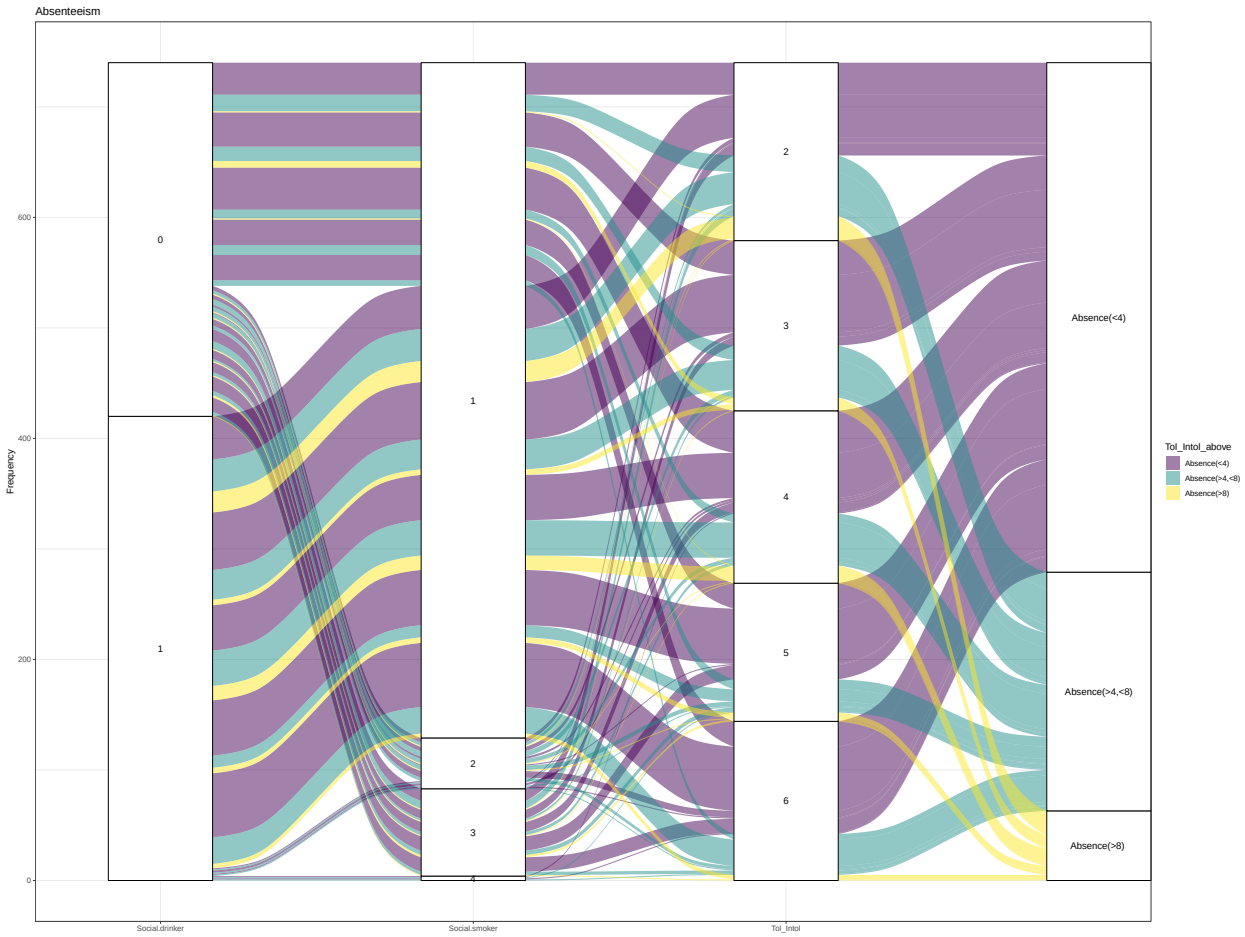
Factored note

```
table_Reason.for.absence<- Abs %>%
  group_by(Social.drinker, Education, Day.of.the.week,Tol_Intol_above) %>%
  count()
```

```
ggplot(table_Reason.for.absence,
       aes(axis2 = Education,
           axis1 = Social.drinker,
           axis3 = Day.of.the.week,
           axis4 = Tol_Intol_above,
           y = n)) +
  geom_alluvium(aes(fill=Tol_Intol_above)) +
  geom_stratum() +
  geom_text(stat = "stratum", aes(label = after_stat(stratum))) +
  scale_x_discrete(limits = c("Social.drinker", "Social.smoker", "Tol_Intol"),
                   expand = c(.1, .2)) +
  scale_fill_viridis_d() +
  labs(title = "Absenteeism",
```

```
       y = "Frequency") +
  theme_bw()
```

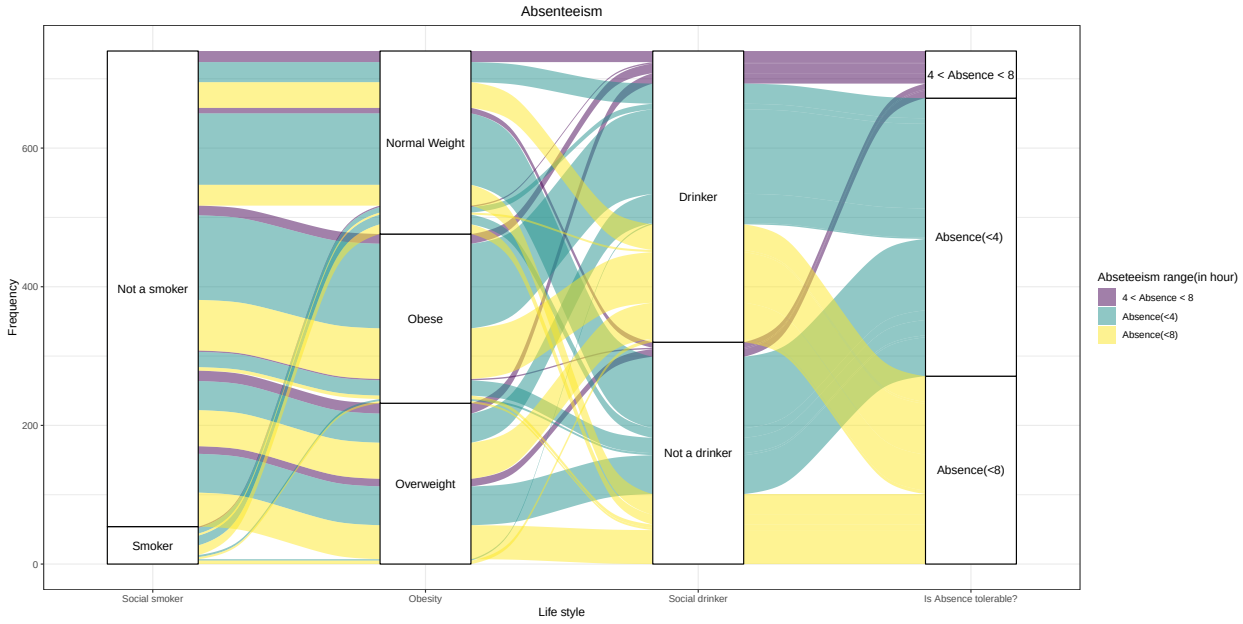
```
## Warning in to_lodes_form(data = data, axes = axis_ind, discern =
## params$discern): Some strata appear at multiple axes.

## Warning in to_lodes_form(data = data, axes = axis_ind, discern =
## params$discern): Some strata appear at multiple axes.

## Warning in to_lodes_form(data = data, axes = axis_ind, discern =
## params$discern): Some strata appear at multiple axes.
```




```
Abs <- Abs %>%
  mutate(Obesity = ifelse(Body.mass.index < 18.5,"Underweight", ifelse(Body.mass.index < 24.9 ,"Normal
```

```
Abs <- Abs %>%
  mutate(Tol_Intol_above = ifelse(Absenteeism.time.in.hours < 4,"Absence(<4) ",ifelse(Absenteeism.time.
table_Obesity<- Abs %>%
  group_by(Social.drinker_1, Obesity, Social.smoker_1,Tol_Intol_above) %>%
  count()
```

```
ggplot(table_Obesity,
       aes(axis1 = Social.smoker_1,
           axis2 = Obesity,
           axis3 = Social.drinker_1,
           axis4 = Tol_Intol_above,
           y = n)) +
  geom_alluvium(aes(fill=Tol_Intol_above)) +
  geom_stratum() +
  geom_text(stat = "stratum", aes(label = after_stat(stratum))) +
  scale_x_discrete(limits = c("Social smoker", "Obesity", "Social drinker","Is Absence tolerable?"),
                   expand = c(.1, .1)) +
  scale_fill_viridis_d() +
  labs(title = "Absenteeism",

       y = "Frequency" ,x ="Life style",fill="Abseteeism range(in hour)") +
  theme_bw()+

    theme(plot.title = element_text(hjust = 0.5))
```



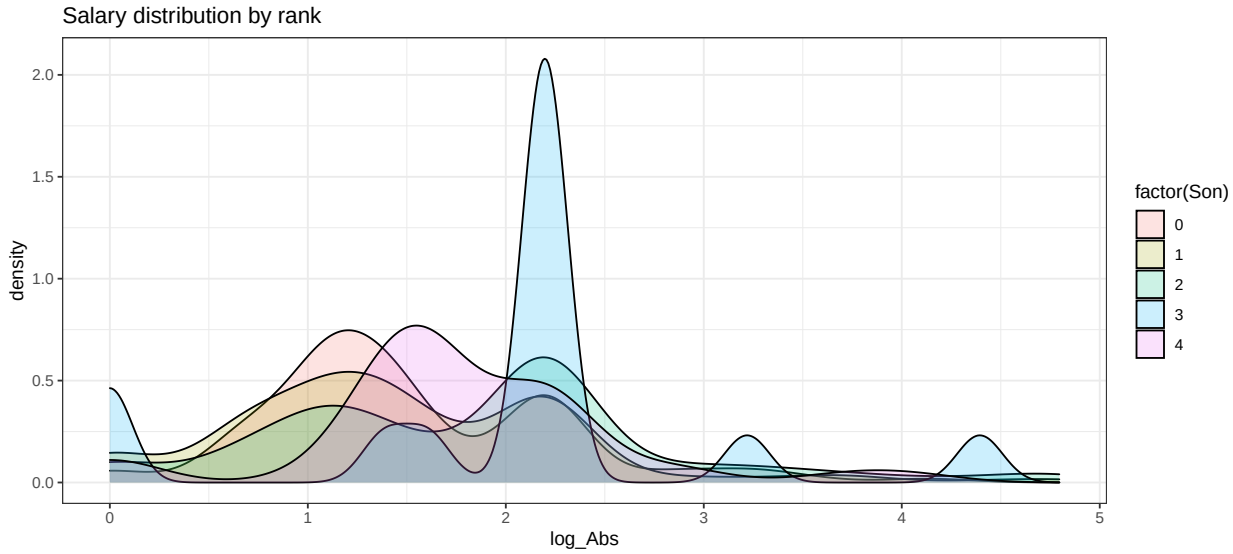

```
Abs%>%
  group_by(Seasons)%>%
  summarise(mean(Absenteeism.time.in.hours))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 4 x 2
##   Seasons `mean(Absenteeism.time.in.hours)`
##     <dbl>                             <dbl>
## 1       1                              7.3
## 2       2                              6
## 3       3                              8.15
## 4       4                              6.35
```

Abs_filtered<-Abs%>% filter(Absenteeism.time.in.hours>0)

```
Abs<-Abs%>%
  mutate(log_Abs = log(Absenteeism.time.in.hours+1))

ggplot(Abs,
       aes(x = log_Abs,
           fill = factor(Son))) +
  geom_density(alpha = 0.2) +
  labs(title = "Salary distribution by rank")+
  theme_bw()
```

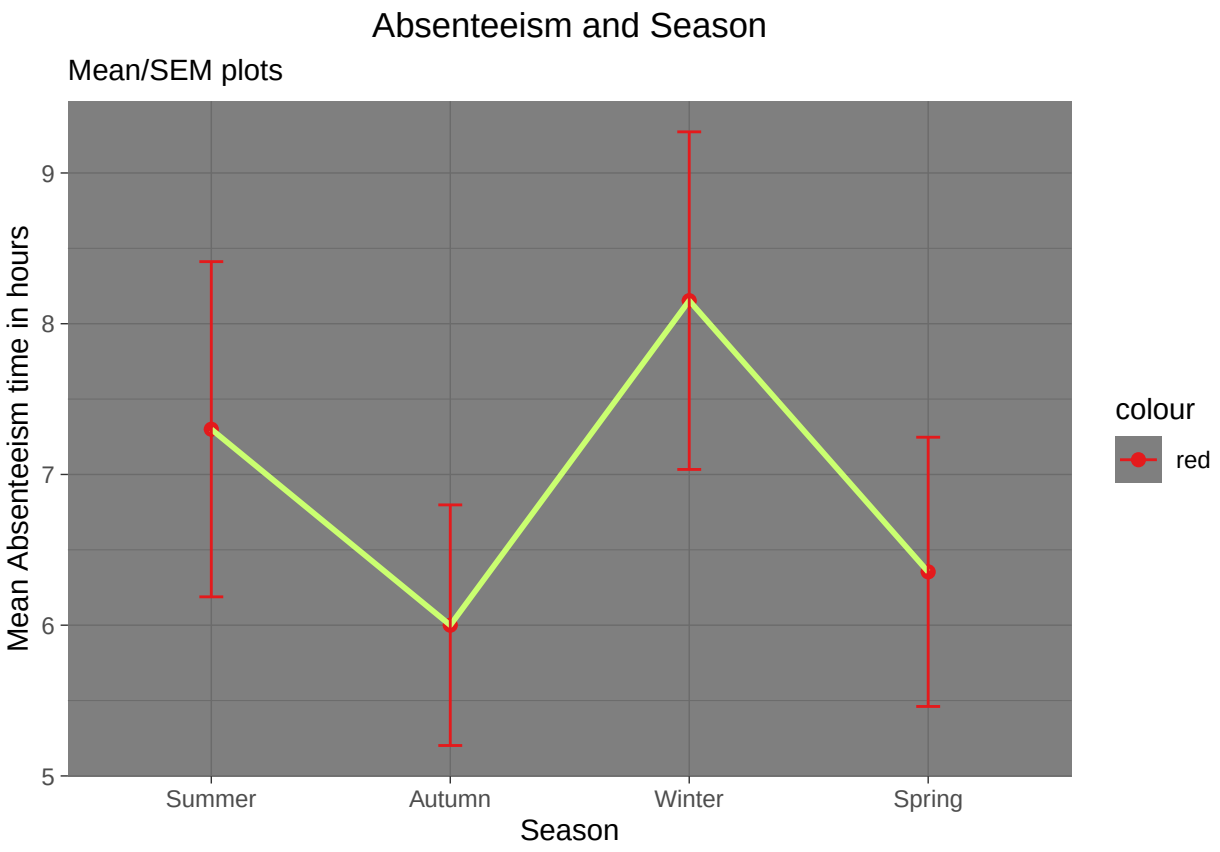



```
library(dplyr)
plotdata_error <- Abs %>%
  group_by(Seasons) %>%
  summarize(n = n(),
          mean = mean(Absenteeism.time.in.hours),
          sd = sd(Absenteeism.time.in.hours),
          se = sd / sqrt(n),
          ci = qt(0.975, df = n - 1) * sd / sqrt(n))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
ggplot(plotdata_error,
       aes(x = factor(Seasons,label=c("Summer", "Autumn", "Winter", "Spring")),
           y = mean,
           group = 1,color="red")) +
  geom_point(size = 2) +
  geom_line(color="darkolivegreen1",size=1) +
  geom_errorbar(aes(ymin = mean - se,
                    ymax = mean + se),
                width = .1)+scale_color_brewer(palette="Set1") +
  xlab("Season")+
  ylab("Mean Absenteeism time in hours")+
```

```
  ggtitle("Absenteeism and Season")+
  labs(subtitle = "Mean/SEM plots")+
  theme_dark()+

    theme(plot.title = element_text(hjust = 0.5))
```




```
Abs%>%
  filter(Reason.for.absence=="26")%>%
  summarise(ID)
```

```
## `summarise()` regrouping output by 'ID' (override with `.groups` argument)
```

```
## # A tibble: 33 x 1
## # Groups:   ID [11]
##       ID
##    <dbl>
##  1     1
##  2     1
##  3     3
##  4     5
##  5     5
##  6     5
##  7     5
##  8     5
```

```
##  9     5
## 10     5
## # ... with 23 more rows
```

```
Unjust_Abs_1<- Abs%>%
  filter(Reason.for.absence=="26")%>%
  group_by(ID)%>%
  summarise(Num_unjust_Abs=n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

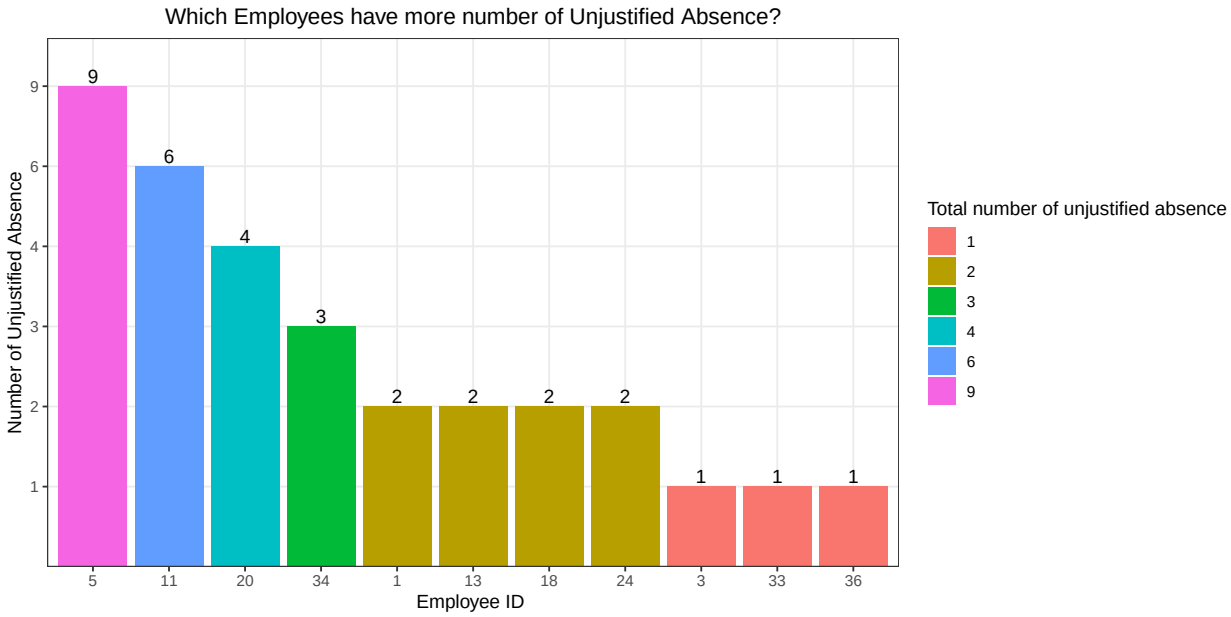
```
Unjust_Abs<-Abs%>%
  filter(Reason.for.absence=="26")%>%
  group_by(ID)%>%
  summarise(Sum_unjust_Abs=sum(Absenteeism.time.in.hours))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

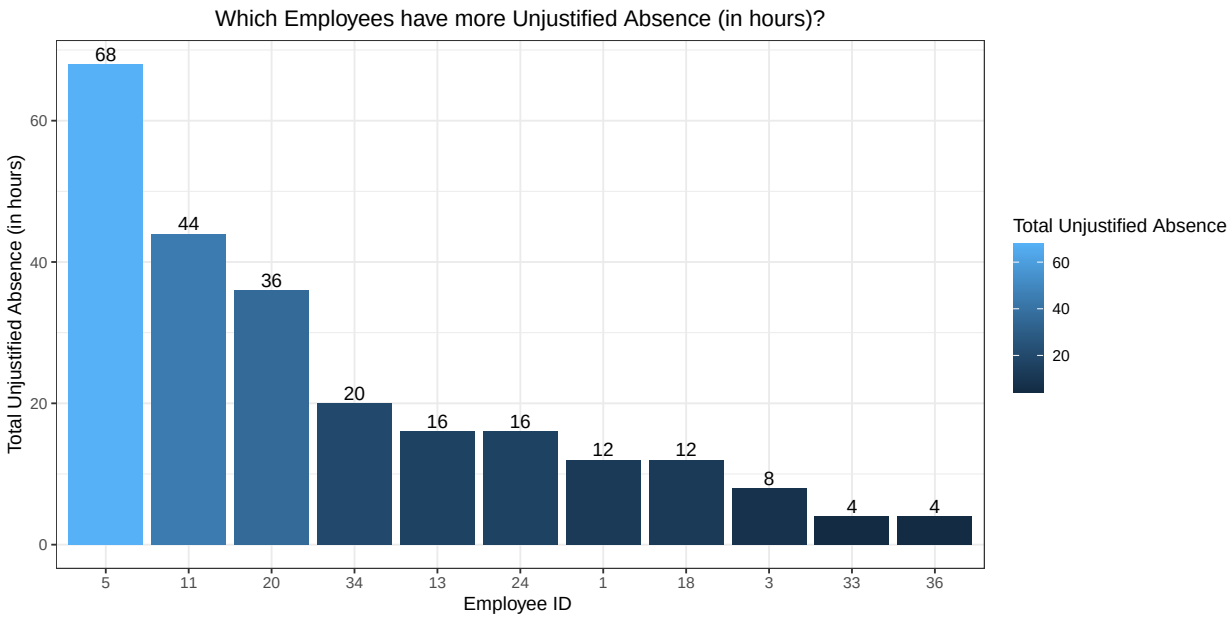
```
Unjust_Abs$Num_unjust_Abs <- factor(Unjust_Abs_1$Num_unjust_Abs)
#Unjust_Abs$ID <- factor(Unjust_Abs$ID)

Unjust_Abs$ID <- factor(Unjust_Abs$ID,
                  levels = Unjust_Abs$ID[order(Unjust_Abs$Num_unjust_Abs, decreasing = TRUE)])
```

,fill=Num_unjust_Abs

```
ggplot(data=Unjust_Abs, aes(x=ID, y=Num_unjust_Abs,fill=Num_unjust_Abs)) +
geom_bar(stat="identity")+
  geom_text(aes(label = Num_unjust_Abs),
            vjust = -0.25) +

  labs(title = "Which Employees have more number of Unjustified Absence?",
       y = "Number of Unjustified Absence", x = "Employee ID ",fill= "Total number of unjustified absen

    theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



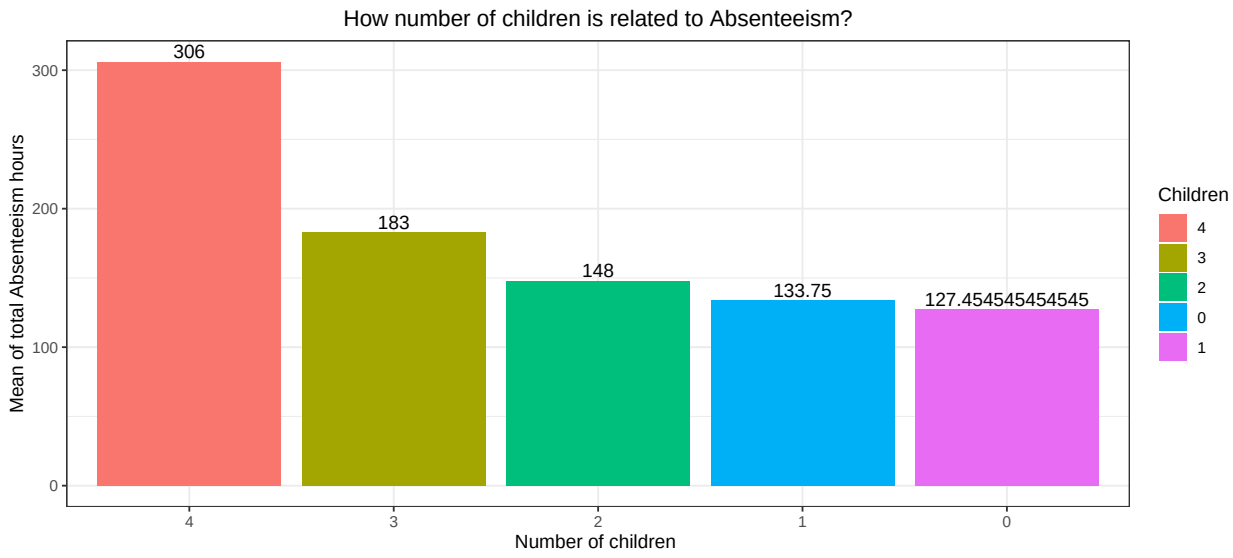

```
Unjust_Abs$ID <- factor(Unjust_Abs$ID,
                 levels = Unjust_Abs$ID[order(Unjust_Abs$Sum_unjust_Abs, decreasing = TRUE)])

ggplot(data=Unjust_Abs, aes(x=ID, y=Sum_unjust_Abs,fill=Sum_unjust_Abs)) +
geom_bar(stat="identity")+
  geom_text(aes(label = Sum_unjust_Abs),
            vjust = -0.25) +

  labs(title = "Which Employees have more Unjustified Absence (in hours)?",
       y = "Total Unjustified Absence (in hours)", x = "Employee ID ", fill = "Total Unjustified Absenc

    theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



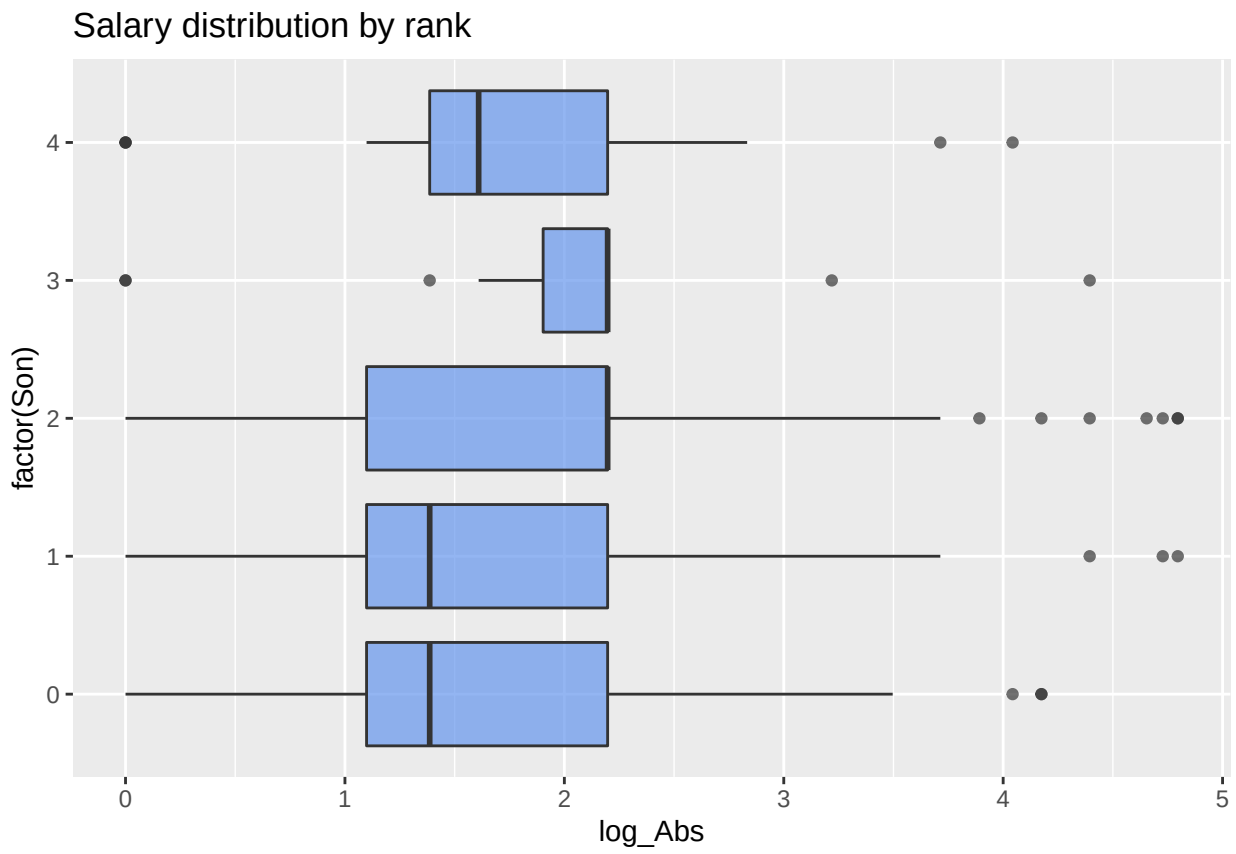
```
Sons<-Abs_single%>%
  group_by(Son)%>%
  summarize(Sum_Abs_mean=mean(Sum_Abs))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
Sons$Son <- factor(Sons$Son,
                   levels = Sons$Son[order(Sons$Sum_Abs_mean, decreasing = TRUE)])

ggplot(data=Sons, aes(x=factor(Son,label=c("4","3","2","1","0")), y=Sum_Abs_mean,fill=Son)) +
geom_bar(stat="identity")+
  geom_text(aes(label = Sum_Abs_mean),
            vjust = -0.25) +

  labs(title = "How number of children is related to Absenteeism?",
       y = "Mean of total Absenteeism hours", x = "Number of children ", fill = "Children")+

    theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



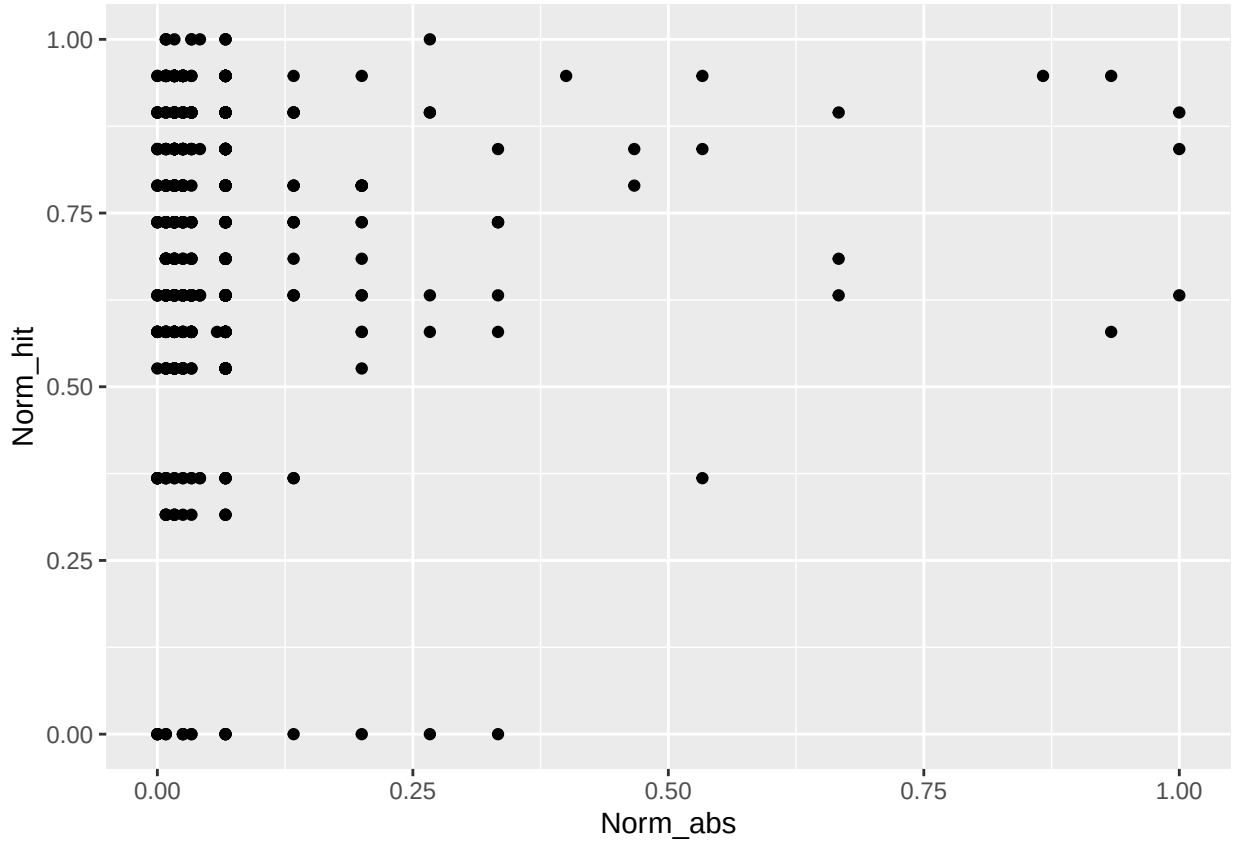

```
Abs_single<-Abs_single%>%
  mutate(log_Num_abs = log(Num_Abs+1))

ggplot(Abs,
       aes(y = factor(Son),
           x = log_Abs)) +
  geom_boxplot(notch = FALSE,
               fill = "cornflowerblue",
               alpha = .7) +
  labs(title = "Salary distribution by rank")
```




```
normalize <- function(x) {
return ((x - min(x)) / (max(x) - min(x)))
}
```

```
Abs$Norm_abs<-normalize(Abs$Absenteeism.time.in.hours)
Abs$Norm_hit<-normalize(Abs$Hit.target)
```

```
ggplot(Abs,
       aes(x = Norm_abs,
           y = Norm_hit)) +
  geom_point()
```




```
Abs%>%
  group_by(Tol_Intol_above)%>%
  summarise(mean(Hit.target))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 3 x 2
##   Tol_Intol_above   `mean(Hit.target)`
##   <chr>                          <dbl>
## 1 "4 < Absence < 8"               93.7
## 2 "Absence(<4) "                  94.7
## 3 "Absence(<8)"                   94.7
```

```
Abs%>%
  group_by(Day.of.the.week)%>%
  summarise(sum(Hit.target))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 5 x 2
##   Day.of.the.week `sum(Hit.target)`
##             <dbl>             <dbl>
## 1               2             15230
## 2               3             14477
```

```
## 3               4             14825
## 4               5             11825
## 5               6             13638
```

```
Abs%>%
  group_by(Month.of.absence)%>%
  summarise(mean(Hit.target))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 13 x 2
##    Month.of.absence `mean(Hit.target)`
##               <dbl>              <dbl>
##  1                0               95
##  2                1               96.4
##  3                2               97.2
##  4                3               97.4
##  5                4               94.6
##  6                5               97.2
##  7                6               95.1
##  8                7               95.0
##  9                8               92.9
## 10                9               86.9
## 11               10               90.3
## 12               11               94.4
## 13               12               95.8
```

```
Abs%>%
  group_by(Seasons)%>%
  summarise(mean(Hit.target))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 4 x 2
##   Seasons `mean(Hit.target)`
##     <dbl>              <dbl>
## 1       1               92.7
## 2       2               96.9
## 3       3               96.0
## 4       4               92.6
```

```
Abs%>%
  group_by(Social.drinker)%>%
  summarise(mean(Hit.target))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)

## # A tibble: 2 x 2
##   Social.drinker `mean(Hit.target)`
##            <dbl>              <dbl>
## 1              0               95.0
## 2              1               94.2
```

```
Abs%>%
  group_by(Service.time)%>%
  summarise(mean(Hit.target))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 18 x 2
##    Service.time `mean(Hit.target)`
##           <dbl>              <dbl>
##  1            1               96.1
##  2            3               94.1
##  3            4               96.2
##  4            6               93.7
##  5            7               95.1
##  6            8               95.2
##  7            9               95.0
##  8           10               94.6
##  9           11               94.8
## 10           12               92.6
## 11           13               93.5
## 12           14               95.0
## 13           15               98.5
## 14           16               94.3
## 15           17               96
## 16           18               94.9
## 17           24               97.5
## 18           29               92.6
```

Which employee has more unjustified absence

https://www.schirrerwalster.lu/unjustified-absence-employer-right-dismiss-worker/

```
Abs%>%
  filter(Reason.for.absence=="26")%>%
  summarise(ID)
```

```
## `summarise()` regrouping output by 'ID' (override with `.groups` argument)
```

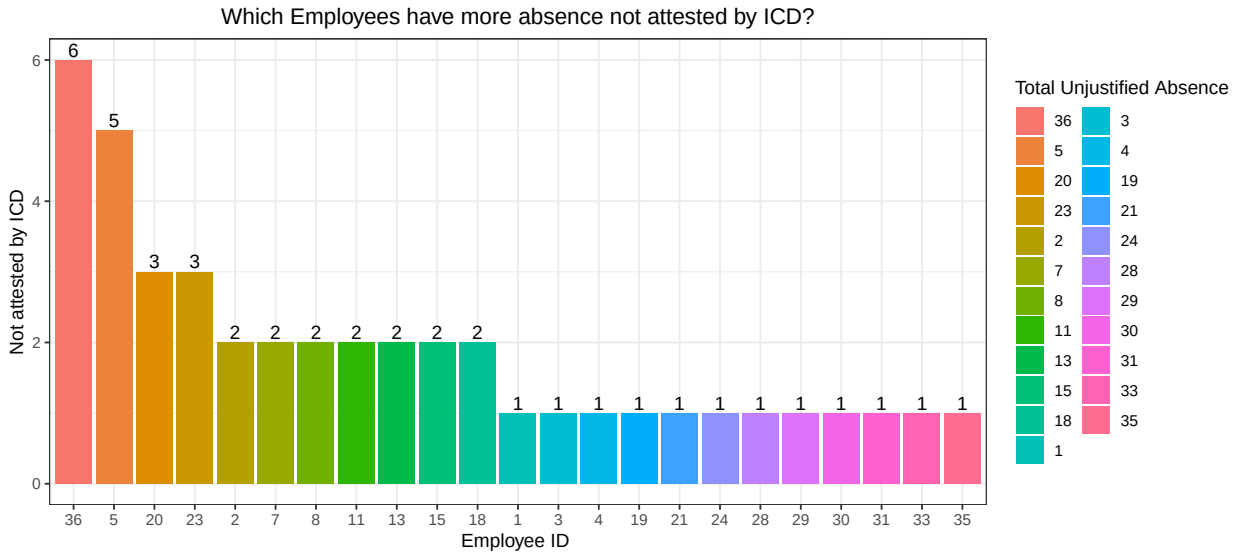
```
## # A tibble: 33 x 1
## # Groups:   ID [11]
##       ID
##    <dbl>
##  1     1
##  2     1
##  3     3
##  4     5
##  5     5
##  6     5
##  7     5
##  8     5
##  9     5
## 10     5
## # ... with 23 more rows
```

'data.frame': 740 obs. of 22 variables: $ ID : num 11 36 3 7 11 3 10 20 14 1 ... $ Reason.for.absence : num 26 0 23 7 23 23 22 23 19 22 ... $ Month.of.absence : num 7 7 7 7 7 7 7 7 7 7 ... $ Day.of.the.week : num 3 3 4 5 5 6 6 6 2 2 ... $ Seasons : num 1 1 1 1 1 1 1 1 1 1 ... $ Transportation.expense : num 289 118 179 279 289 179 361 260 155 235 ... $ Distance.from.Residence.to.Work: num 36 13 51 5 36 51 52 50 12 11 ... $ Service.time : num 13 18 18 14 13 18 3 11 14 14 ... $ Age : num 33 50 38 39 33 38 28 36 34 37 ... $ Work.load.Average.day : num 240 240 240 240 240 ... $ Hit.target : num 97 97 97 97 97 97 97 97 97 97 ... $ Disciplinary.failure : num 0 1 0 0 0 0 0 0 0 0 ... $ Education : num 1 1 1 1 1 1 1 1 1 3 ... $ Son : num 2 1 0 2 2 0 1 4 2 1 ... $ Social.drinker : num 1 1 1 1 1 1 1 1 1 0 ... $ Social.smoker : num 0 0 0 1 0 0 0 0 0 0 ... $ Pet : num 1 0 0 0 1 0 4 0 0 1 ... $ Weight : num 90 98 89 68 90 89 80 65 95 88 ... $ Height : num 172 178 170 168 172 170 172 168 196 172 ... $ Body.mass.index : num 30 31 31 24 30 31 27 23 25 29 ... $ Absenteeism.time.in.hours : num 4 0 2 4 2 2 8 4 40 8 ... $ icd_attested

```
Not_ICD<-Abs%>%

  filter(Reason.for.absence=="0")%>%
 group_by(ID)%>%
  summarise(Not_ICD = n())
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
Not_ICD$ID <- factor(Not_ICD$ID,
                     levels = Not_ICD$ID[order(Not_ICD$Not_ICD, decreasing = TRUE)])

ggplot(data=Not_ICD, aes(x=ID, y=Not_ICD,fill=ID)) +
geom_bar(stat="identity")+
  geom_text(aes(label = Not_ICD),
            vjust = -0.25) +

  labs(title = "Which Employees have more absence not attested by ICD?",
       y = "Not attested by ICD", x = "Employee ID ", fill = "Total Unjustified Absence")+

    theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



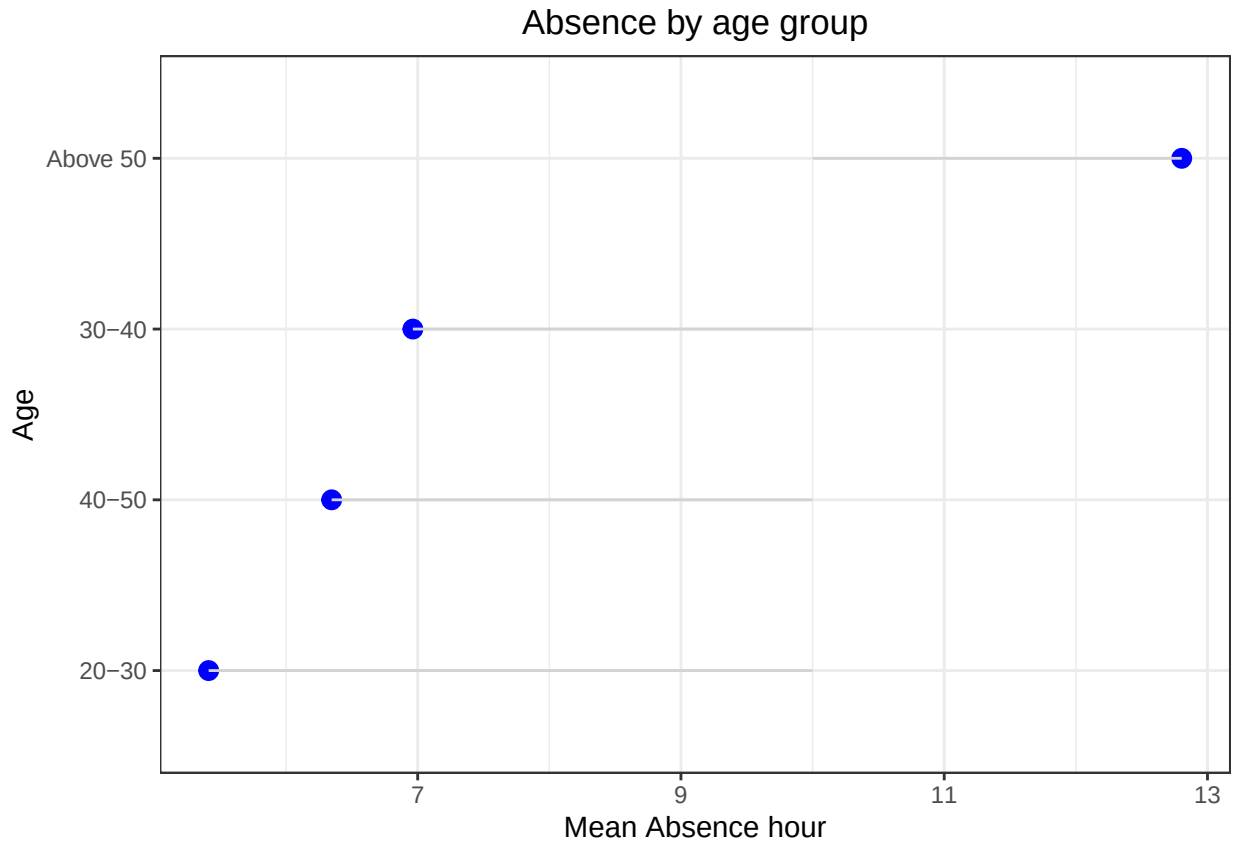
Abs <- Abs%>% mutate(Age_grp = ifelse(Age<30,"25-30",ifelse(Age<35,"30-35",ifelse(Age<40,"35-40",ifelse(Age<45,"40-45",ifelse(Age<50,"45-50",ifelse(Age<55,"50-55","Above 55")))))))

```
Abs <- Abs%>%
  mutate(Age_grp = ifelse(Age<30,"20-30",ifelse(Age<40,"30-40",ifelse(Age<50,"40-50","Above 50"))))
```

```
Age_grp_abs<- Abs%>%
  group_by(Age_grp)%>%
  summarise(Age_Abs = mean(Absenteeism.time.in.hours))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
ggplot(Age_grp_abs,
       aes(x=Age_Abs,
           y=reorder(Age_grp, Age_Abs))) +
  geom_point(color="blue",
             size = 3)  +
  geom_segment(aes(x = 10,
                xend = Age_Abs,
                y = reorder(Age_grp, Age_Abs),
                yend = reorder(Age_grp, Age_Abs)),
                color = "lightgrey")+
  labs (x = "Mean Absence hour",
        y = "Age",
        title = "Absence by age group") +
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



1 : 'infectious parasitic', 2 : 'neoplasms', 3 : 'blood blood-forming immune', 4 : 'endocrine nutritional

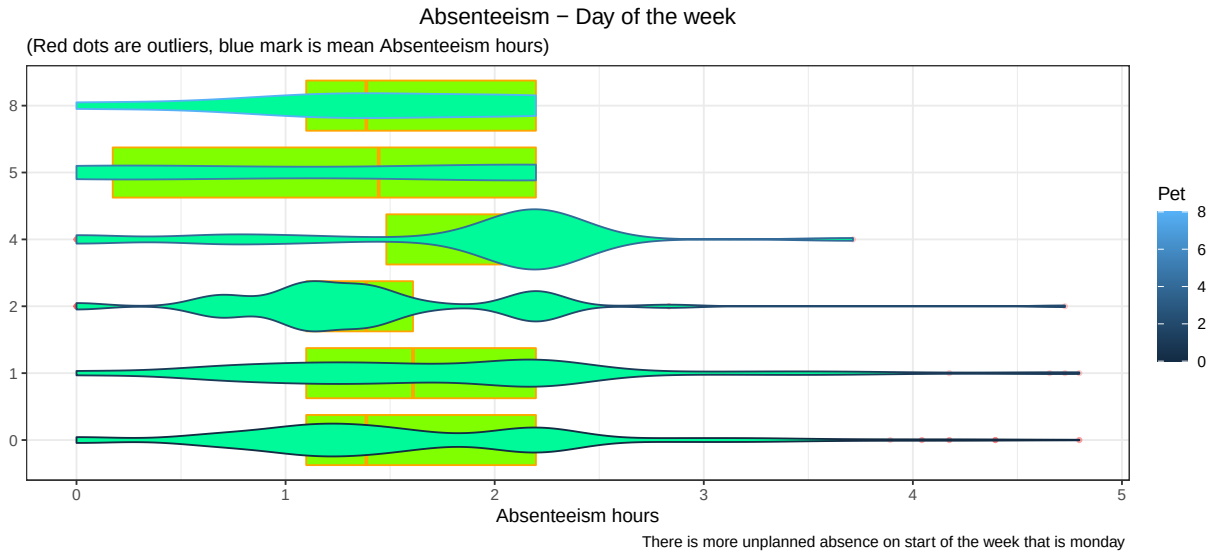
metabolic', 5 : 'mental behavioral', 6 : 'nervous', 7 : 'eye adnexa', 8 : 'ear mastoid', 9 : 'circulatory', 10 : 'respiratory', 11 : 'digestive', 12 : 'skin subcutaneous', 13 : 'musculoskeletal connective', 14 : 'genitourinary', 15 : 'pregnancy childbirth puerperium', 16 : 'perinatal', 17 : 'congenital malformations deformations chromosomal', 18 : 'clinical laboratory', 19 : 'injury poisoning', 20 : 'morbidity mortality', 21 : 'health status services', 22 : 'patient', 23 : 'medical', 24 : 'blood', 25 : 'laboratory', 26 : 'unjustified', 27 : 'physiotherapy', 28 : 'dental'

```
Abs%>%
  group_by(Pet)%>%
  summarise(mean(Absenteeism.time.in.hours))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 6 x 2
##     Pet `mean(Absenteeism.time.in.hours)`
##   <dbl>                             <dbl>
## 1     0                              6.83
## 2     1                              9
## 3     2                              4.75
## 4     4                              7.06
## 5     5                              4.17
## 6     8                              4.25
```

```
ggplot(data=Abs,

       aes(y= factor(Pet)

           ,x=log_Abs,color =Pet )) +
  geom_boxplot(fill = "chartreuse1", colour = "orange",outlier.color = "red",
               outlier.alpha = 0.2,
               outlier.size = 1)+ stat_summary(fun=mean, geom="point", shape=3, size=0.8, color="blue")+
  geom_violin(fill = "mediumspringgreen")+

  labs(title="Absenteeism - Day of the week",subtitle = "(Red dots are outliers, blue mark is mean Absen
       x = "Absenteeism hours",y="")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```




```
Abs_single<- Abs_single%>%
  mutate(Pet_Yes_No = ifelse(Pet=="0","No pet","Have pets"))
```

```
Abs_single%>%
  group_by(Pet_Yes_No)%>%
  summarise(mean(Sum_Abs))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
## # A tibble: 2 x 2
##   Pet_Yes_No `mean(Sum_Abs)`
##   <chr>                <dbl>
## 1 Have pets             117.
## 2 No pet                165.
```

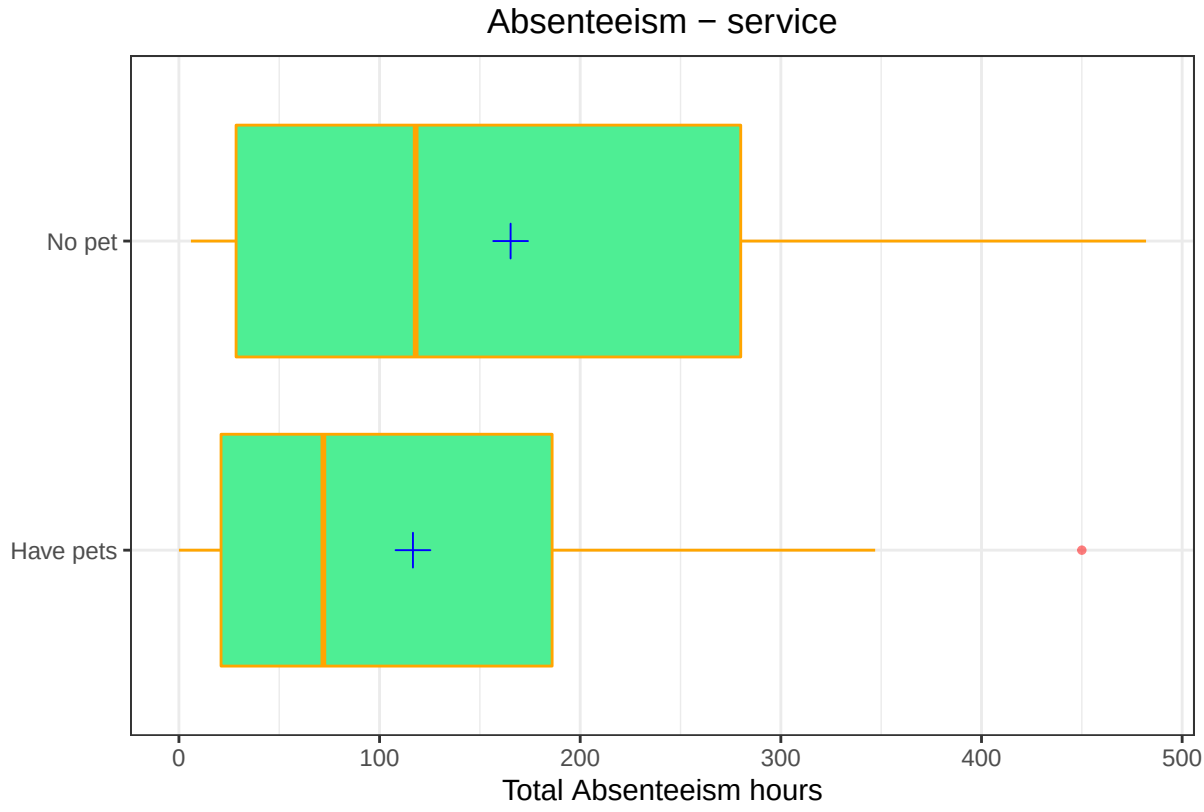
```
Abs_single<- Abs_single%>%
  mutate(Son_Yes_No = ifelse(Son=="0","No children","Have children"))
```

```
Abs_single%>%
  group_by(Son_Yes_No)%>%
  summarise(mean(Sum_Abs))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

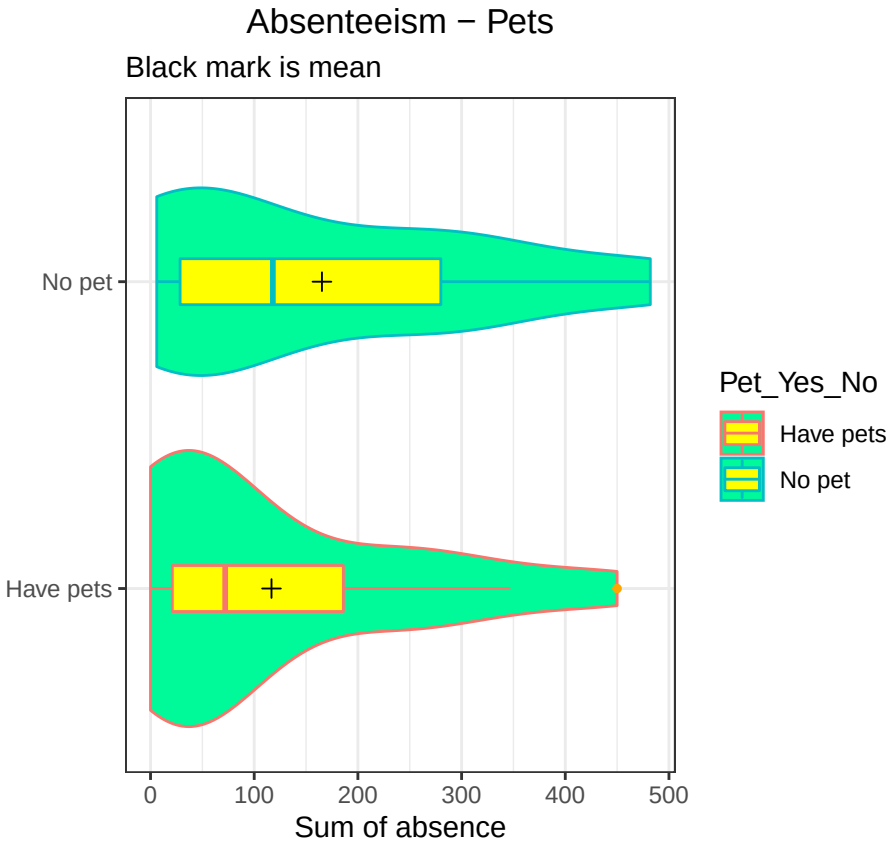
```
## # A tibble: 2 x 2
##   Son_Yes_No    `mean(Sum_Abs)`
##   <chr>                   <dbl>
## 1 Have children            147.
## 2 No children              134.
```

```
ggplot(data=Abs_single,

       aes(y= Pet_Yes_No

           ,x=Sum_Abs,color =Pet_Yes_No ))+
  geom_boxplot(fill = "seagreen2", colour = "orange",outlier.color = "red",
               outlier.alpha = 0.5,
               outlier.size = 1)+stat_summary(fun=mean, geom="point", shape=3, size=4, color="blue")+


  scale_x_continuous(labels = scales::comma)+

  labs(title="Absenteeism - service", caption = "People with more than 15 years of service have high mea
       x = "Total Absenteeism hours",y="")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```




```
ggplot(data=Abs_single,

       aes(y=Pet_Yes_No

           ,x=Sum_Abs,color =Pet_Yes_No ))+
  geom_violin(fill = "mediumspringgreen") +
  geom_boxplot(width = .15,
               fill = "yellow",
```

```
              outlier.color = "orange",
              outlier.size = 1) + stat_summary(fun=mean, geom="point", shape=3, size=2, color="black")+
  labs(title="Absenteeism - Pets",subtitle ="Black mark is mean",
       y = "",x="Sum of absence",fill = "Do the employee have pets")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



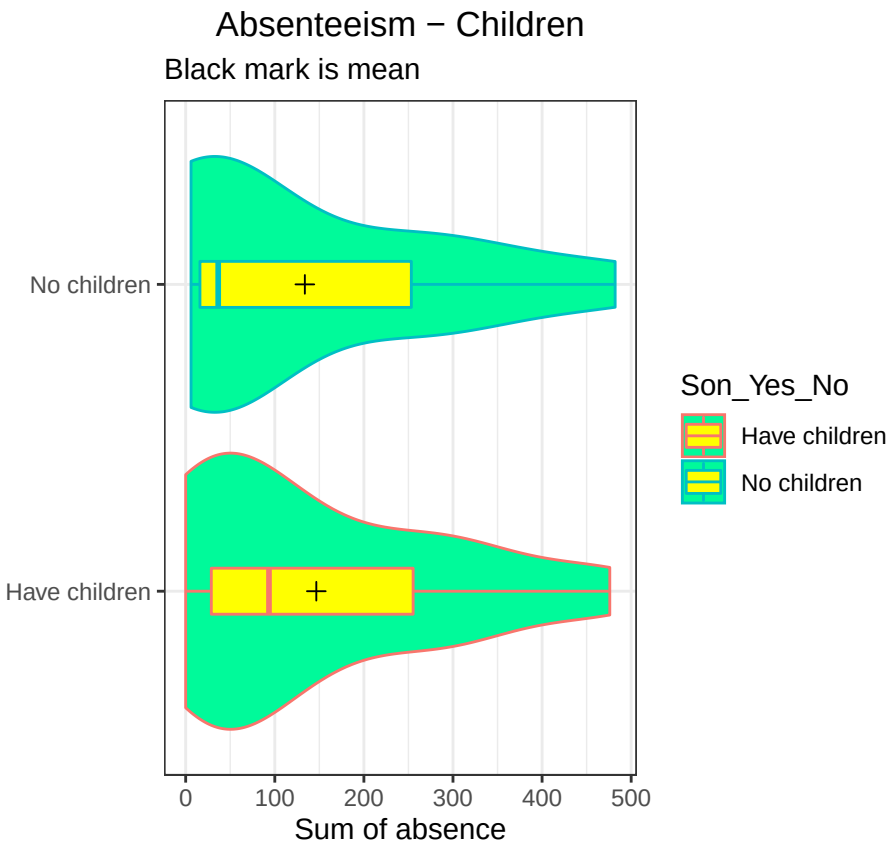
ggplot(data=Unjust_Abs, aes(x=ID, y=Num_unjust_Abs,fill=Num_unjust_Abs)) + geom_bar(stat="identity")+ geom_text(aes(label = Num_unjust_Abs), vjust = -0.25) +

labs(title = "Which Employees have more number of Unjustified Absence?", y = "Number of Unjustified Absence", x = "Employee ID",fill= "Total number of unjustified absence")+

```
theme_bw()+
theme(plot.title = element_text(hjust = 0.5))
```

```
ggplot(data=Abs_single,

       aes(y=Son_Yes_No

           ,x=Sum_Abs,color =Son_Yes_No ))+
  geom_violin(fill = "mediumspringgreen") +
  geom_boxplot(width = .15,
               fill = "yellow",
               outlier.color = "orange",
               outlier.size = 1) + stat_summary(fun=mean, geom="point", shape=3, size=2, color="black")+
```

```
  labs(title="Absenteeism - Children",subtitle ="Black mark is mean",
       y = "",x="Sum of absence",fill = "Do the employee have children")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```




```
plotdata_5<-Abs_single%>%
  group_by(Son)%>%
  summarise(mean_abs_edu= mean(Sum_Abs))
```

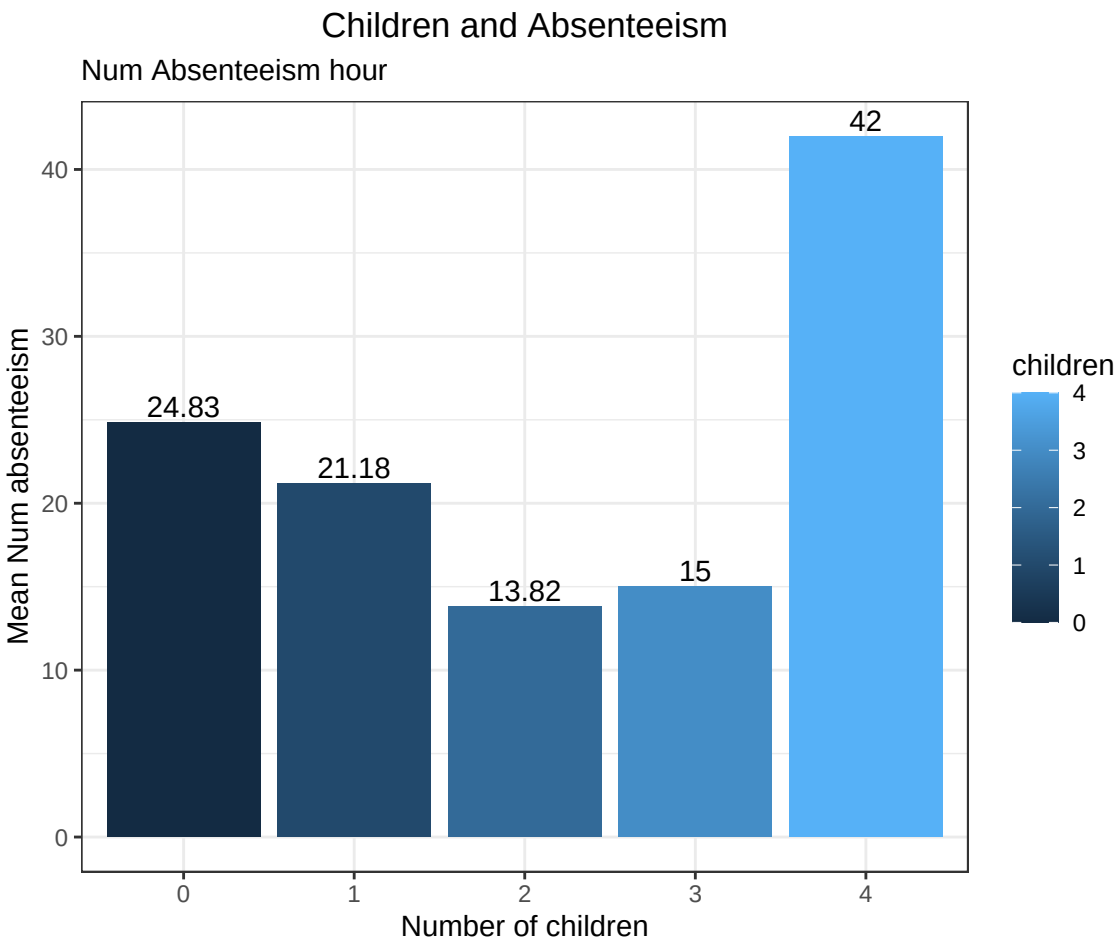
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
plotdata_7<-Abs_single%>%
  group_by(Son)%>%
  summarise(mean_abs_edu= mean(Num_Abs))
```

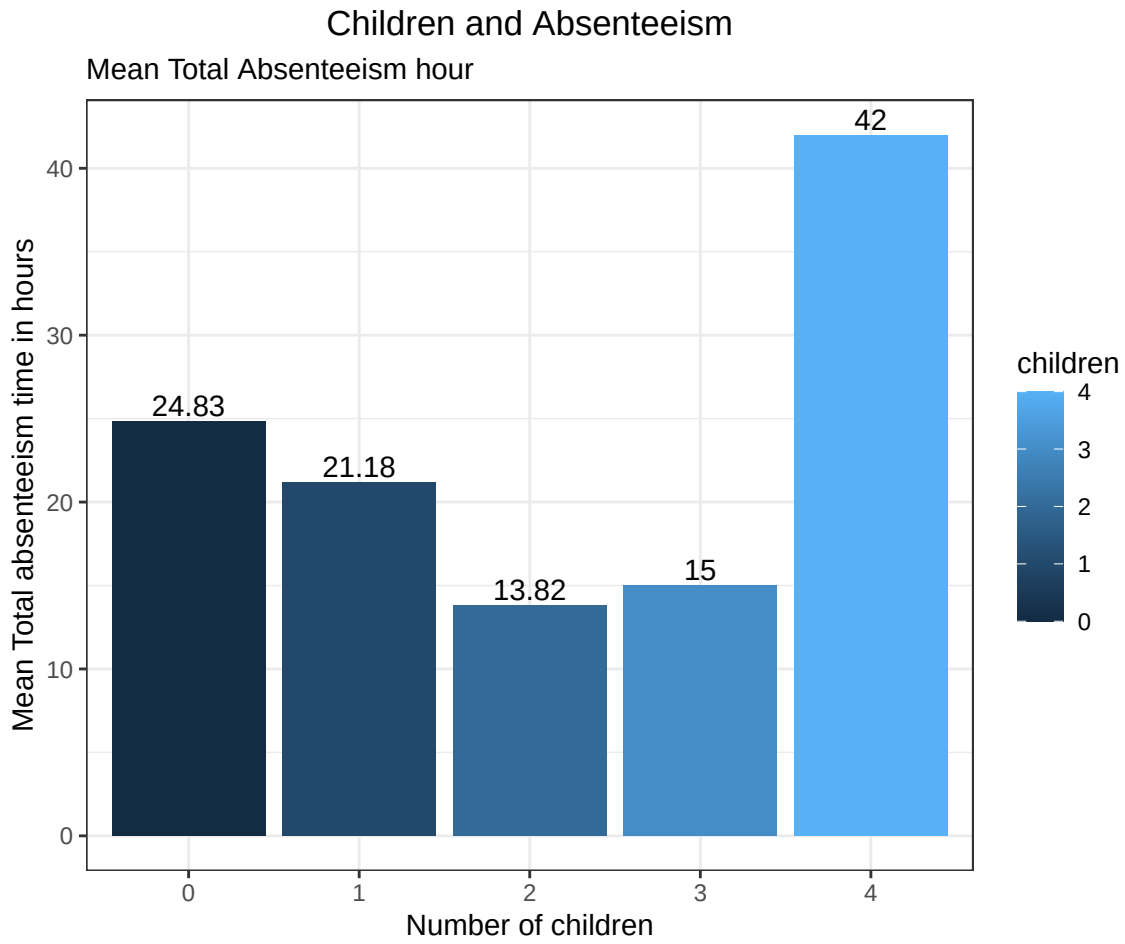
```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
ggplot(plotdata_7,
       aes(x = factor(Son),
                      y = mean_abs_edu,fill=Son)) +
  geom_bar(stat = "identity" ) +
 geom_text( aes(label = round(mean_abs_edu,digits = 2)),
            vjust = -0.25)+
```

```
  labs(title = "Children and Absenteeism",
       subtitle = "Num Absenteeism hour",
       x = "Number of children",
       y = "Mean Num absenteeism",fill="children")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```



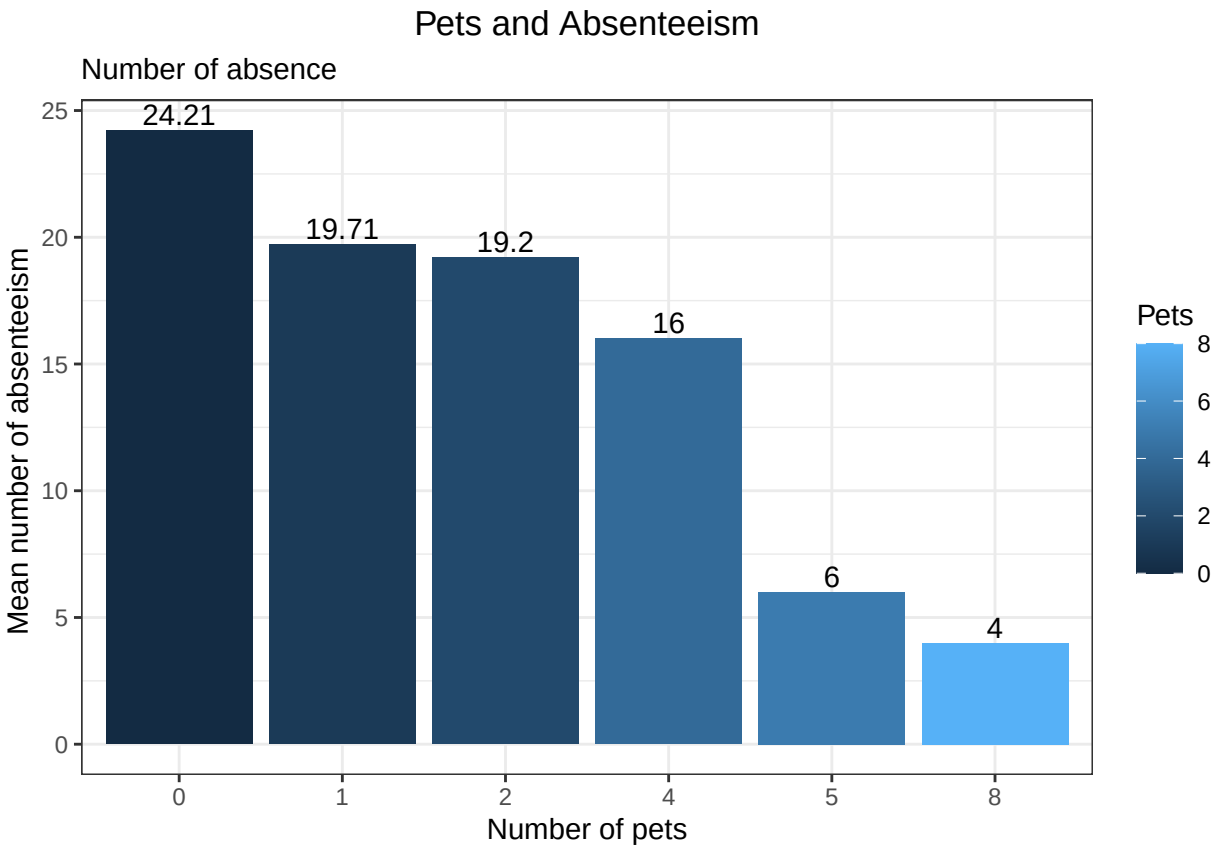

```
ggplot(plotdata_7,
       aes(x = factor(Son),
                      y = mean_abs_edu,fill=Son)) +
  geom_bar(stat = "identity" ) +
 geom_text( aes(label = round(mean_abs_edu,digits = 2)),
            vjust = -0.25)+
  labs(title = "Children and Absenteeism",
       subtitle = "Mean Total Absenteeism hour",
       x = "Number of children",
       y = "Mean Total absenteeism time in hours",fill="children")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```




```
plotdata_6<-Abs_single%>%
  group_by(Pet)%>%
  summarise(mean_abs_edu= mean(Num_Abs))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
ggplot(plotdata_6,
       aes(x = factor(Pet),
                     y = mean_abs_edu,fill=Pet)) +
  geom_bar(stat = "identity" ) +
 geom_text( aes(label = round(mean_abs_edu,digits = 2)),
            vjust = -0.25)+
  labs(title = "Pets and Absenteeism",
       subtitle = "Number of absence ",
       x = "Number of pets",
       y = "Mean number of absenteeism ",fill="Pets")+
  theme_bw()+
    theme(plot.title = element_text(hjust = 0.5))
```


Pets and Absenteeism
Number of absence
Mean number of absenteeism
25
20
15
10
5
0
24.21
19.71
19.2
16
6
4
0
1
2
4
5
8
Number of pets
Pets
8
6
4
2
0