Tut10

Ananya_Gupta

10/25/2021

```r
absent <- read.csv("../data/Absenteeism_at_work.csv",sep=";") %>%
  as_tibble()

absent <- absent %>%
  mutate(Day.of.the.week = recode(Day.of.the.week,
                                  `2`="Mon", `3`="Tue",`4`="Wed", `5`="Thu", `6`="Fri"),
         Seasons=recode(Seasons,
                        `1`="Summer", `2`="Autumn",`3`="Winter", `4`="Spring"))%>%
  filter(Disciplinary.failure!=1)

#Qn1
abs<-absent %>%
  group_by(ID) %>%
  summarise(tot=sum(Absenteeism.time.in.hours))

ordered<-abs$ID[order(sapply(1:36, function(x) abs$tot[x]))]

absent %>%
  group_by(ID) %>%
  mutate(cumulative=cumsum(Absenteeism.time.in.hours))%>%
  ggplot()+
  geom_col(aes(x=Absenteeism.time.in.hours,
               y=factor(ID,levels=ordered),
               fill=Absenteeism.time.in.hours))+
  geom_tile(mapping=aes(x=cumulative,
                        y=as.factor(ID)),
            color="white")+
  labs(x="Hours", y="Courier ID", title="Accumulated and Individual Absence Times")+
  scale_fill_continuous(name="Time (hrs)")
```
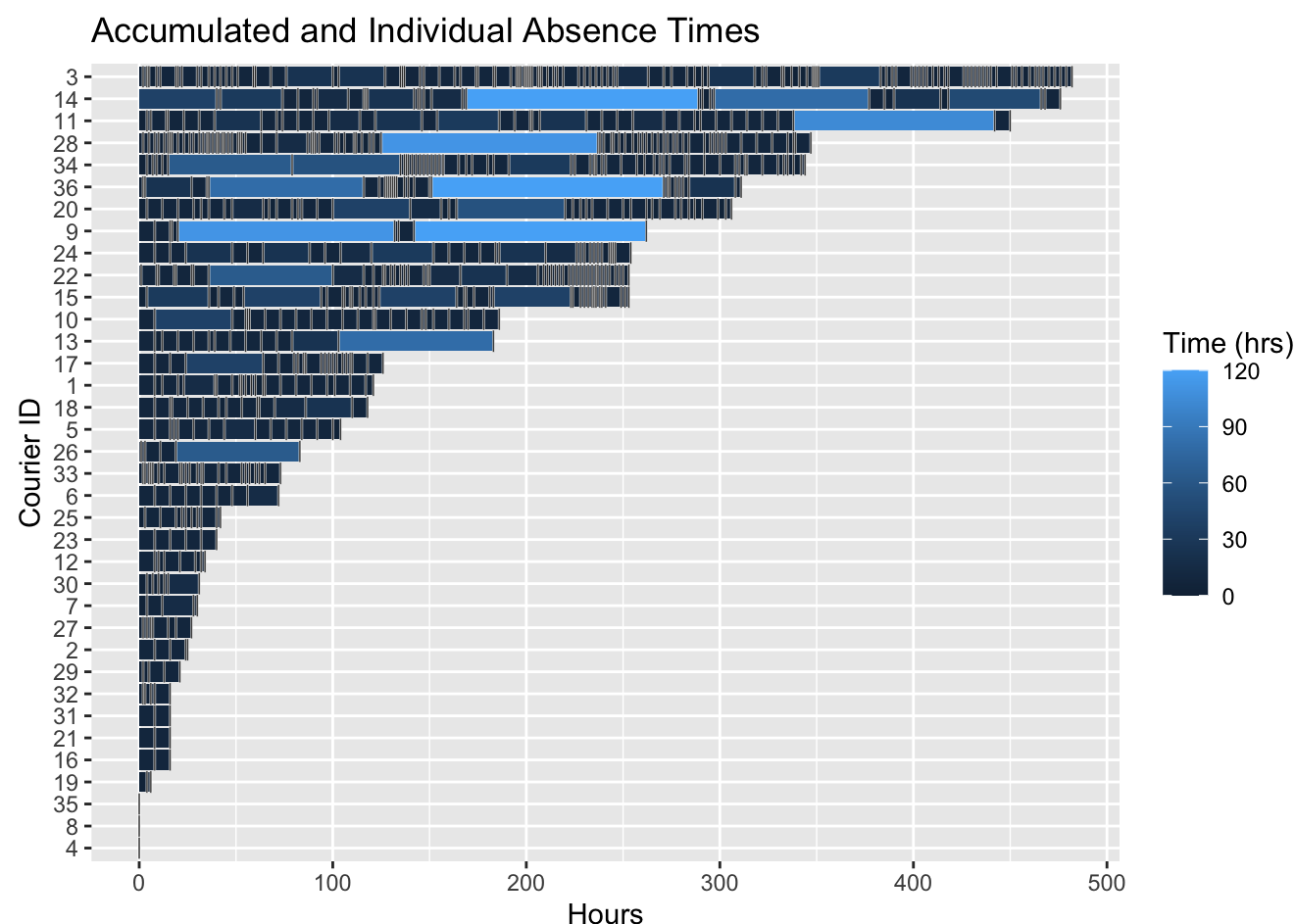

Accumulated and Individual Absence Times

```r
#Qn2
absence_props<-absent %>%
  group_by(Seasons,Day.of.the.week) %>%
  count() %>%
  group_by(Seasons) %>%
  mutate(prop=n/sum(n)) %>%
  ungroup()

absence_props %>%
  mutate(day=factor(Day.of.the.week,levels=c("Mon","Tue","Wed","Thu","Fri")))%>%
  ggplot()+
  geom_col(mapping=aes(x=Seasons,y=prop,fill=day))+
  labs(x="Season", y="Proportion",title="Proportion of Absence per day")+
  scale_fill_manual(values=c("Mon"="lightcoral",
                             "Tue"="gold1",
                             "Wed"="seagreen3",
                             "Thu"="tan1",
                             "Fri"="lightskyblue2"),
                    name="Day of the week",
                    labels=c("Monday","Tuesday","Wednesday","Thursday","Friday"))+
  scale_y_continuous(breaks = seq(0, 1, by=0.05))+
  theme(plot.title=element_text(hjust=0.5))
```


Proportion of Absence per day
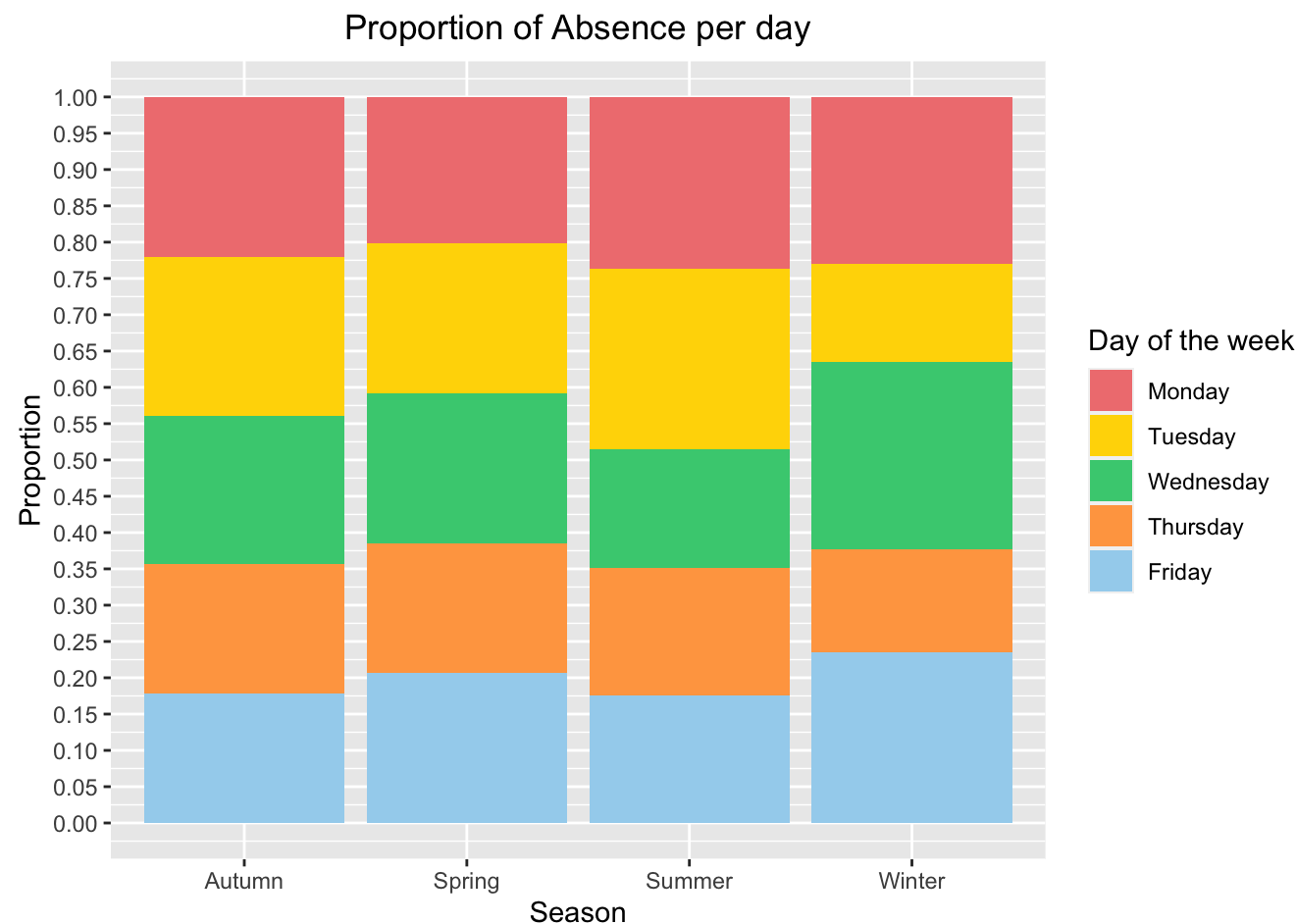
```r
#Qn3
#Trial 1
absent %>%
  mutate(day=factor(Day.of.the.week,levels=c("Mon","Tue","Wed","Thu","Fri")))%>%
  ggplot()+
  geom_bar(aes(x=day,fill=day))+
  labs(x="Day of the week",y="Count",title="Pattern of Absences of each courier per day")+
  facet_wrap(~ID)+
  theme(plot.title=element_text(hjust=0.5),
        axis.text.x=element_text(angle=90, hjust=0.5,vjust=0.5))
```


Pattern of Absences of each courier per day
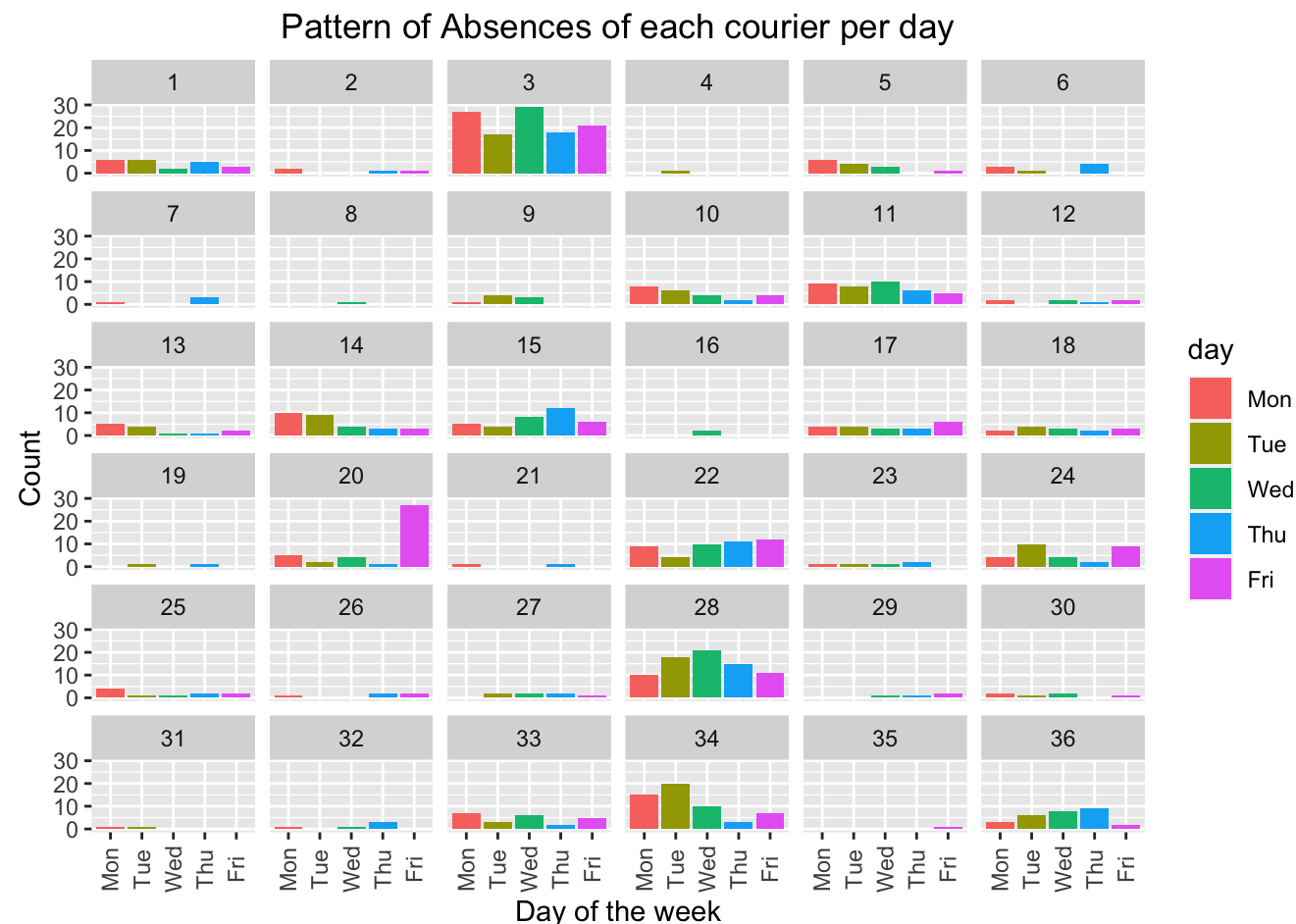
```r
#Trial 2
absent_line<-absent %>%
  select(1,4,5,21) %>%
  group_by(ID,Day.of.the.week)%>%
  count()%>%
  group_by(ID)%>%
  mutate(prop=n/sum(n))%>%
  mutate(cum=cumsum(prop))

absent_line %>%
  ggplot()+
  geom_point(aes(x=Day.of.the.week,y=cum,color=Day.of.the.week))+
  facet_wrap(~ID)+
  labs(x="Day of the week",
       y="Cumulative Proportion",
       title="Proportion pattern of Absences of each courier")+
  theme(plot.title=element_text(hjust=0.5),
        axis.text.x=element_text(angle=90, hjust=0.5,vjust=0.5))+
  scale_color_manual(values=c("Mon"="lightcoral","Tue"="gold1","Wed"="seagreen3","Thu"="tan1","Fri"="lightskyblue
2"),
                     name="Day of the week",
                     labels=c("Monday","Tuesday","Wednesday","Thursday","Friday"))
```


Proportion pattern of Absences of each courier