DSA2101 Tutorial 10

Waseem

11/4/2021

Visualisation on tutorial 5 courier absenteeism

```
library(readr)
library(tidyverse)
```

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

```
## v ggplot2 3.3.5     v dplyr   1.0.7
## v tibble  3.1.4     v stringr 1.4.0
## v tidyr   1.1.3     v forcats 0.5.1
## v purrr   0.3.4
```

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

```
absent <- read_delim("data/Absenteeism_at_work.csv")
```

```
## Rows: 740 Columns: 21
```

```
## -- Column specification --------------------------------------------------------
## Delimiter: ";"
## dbl (21): ID, Reason for absence, Month of absence, Day of the week, Seasons...
```

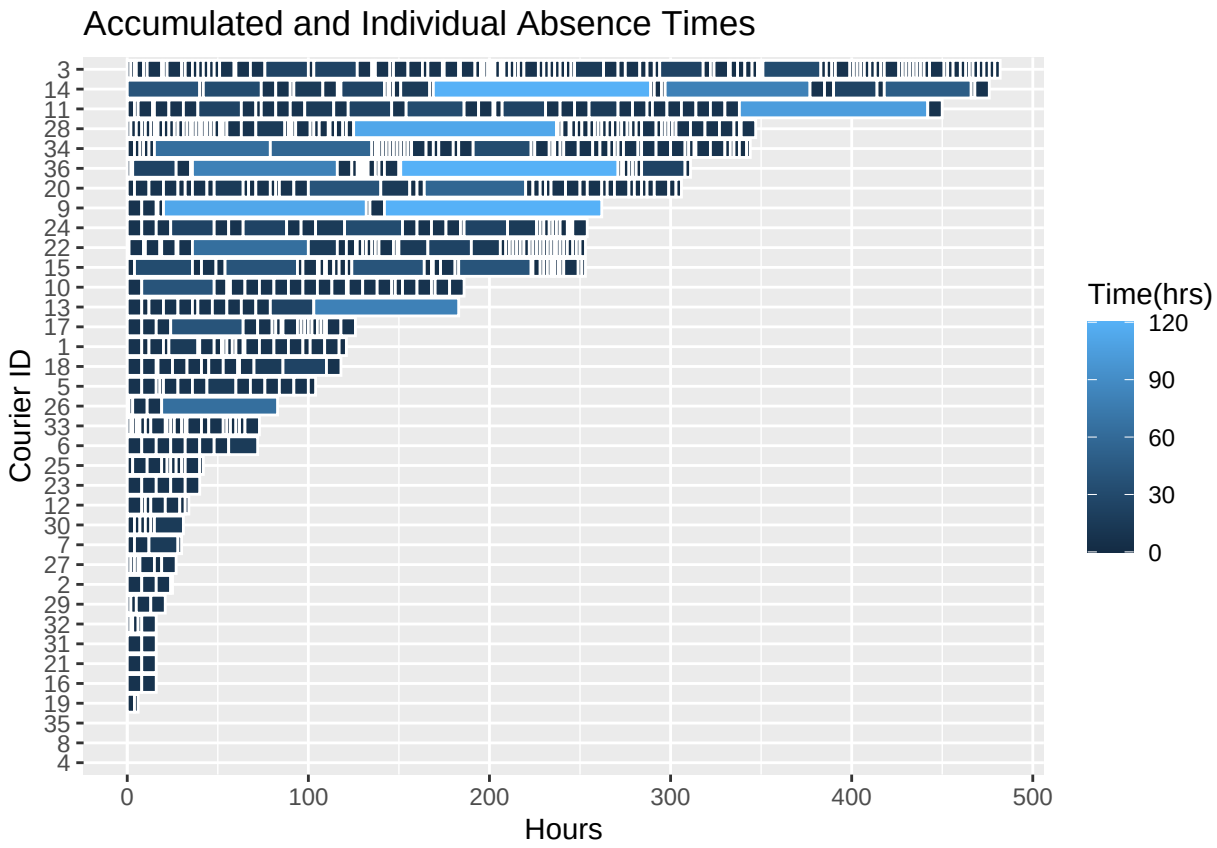
```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

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

Visualisation on Proportion of Absences by Season and Day

```
absent_plot <- absent %>%
  group_by(ID) %>%
  mutate(count=sum(`Absenteeism time in hours`)) %>%
  arrange(desc(count))

ggplot(absent_plot, aes(x=reorder(ID,count),y=`Absenteeism time in hours`,fill=`Absenteeism time in hou
  geom_col(position="stack",color="white")+
  coord_flip()+
  labs(title='Accumulated and Individual Absence Times' ,fill="Time(hrs)" ,x="Courier ID",y="Hours")
```

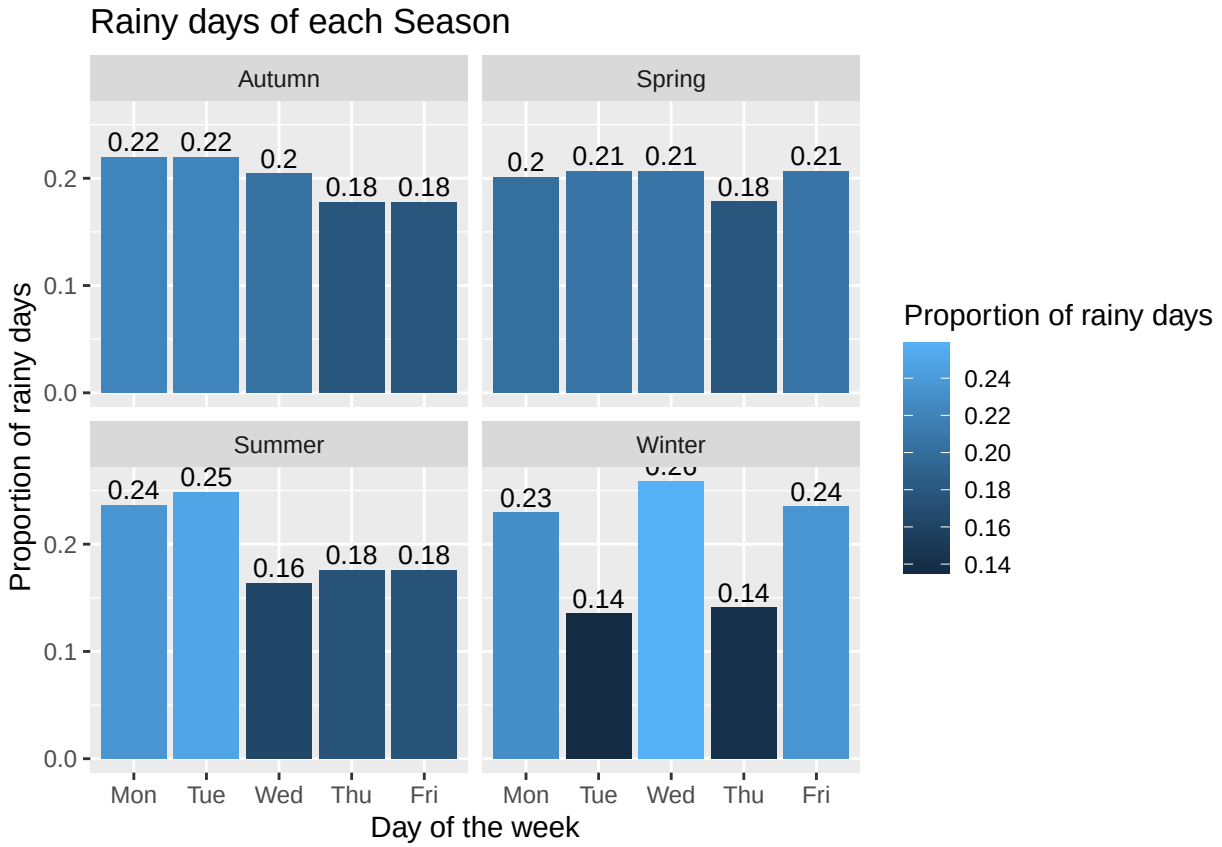


Plot to show the proportion of absences on each day within a season

```
absence_props<-absent %>%
   group_by(Seasons,`Day of the week`) %>%
   count() %>%
   group_by(Seasons) %>%
   mutate(prop= n/sum(n)) %>%
   ungroup()%>%
  mutate(`Day of the week` = fct_relevel(`Day of the week`,
            "Mon", "Tue", "Wed", "Thu", "Fri"))

level <- factor(absence_props$`Day of the week`, levels = c("Mon", "Tue", "Wed", "Thu", "Fri"))

ggplot(absence_props, aes(x= `Day of the week`, y = prop, fill = prop)) +
  geom_col() +
  facet_wrap(~ Seasons) +
  labs(title = "Rainy days of each Season", fill = "Proportion of rainy days",y="Proportion of rainy da
    geom_text(aes(label=round(prop,2)), vjust=-0.3, color="black", size=3.5)
```



Comparing to part 1, where the accumulated and individual absence times were represented, we could adopt a similar method to compare the days instead of the hours that they are absent.