S2208 MATH8050 Data Analysis - Section 001: Homework 2 Due on 09/14/22

Adithya Ravi, C09059838

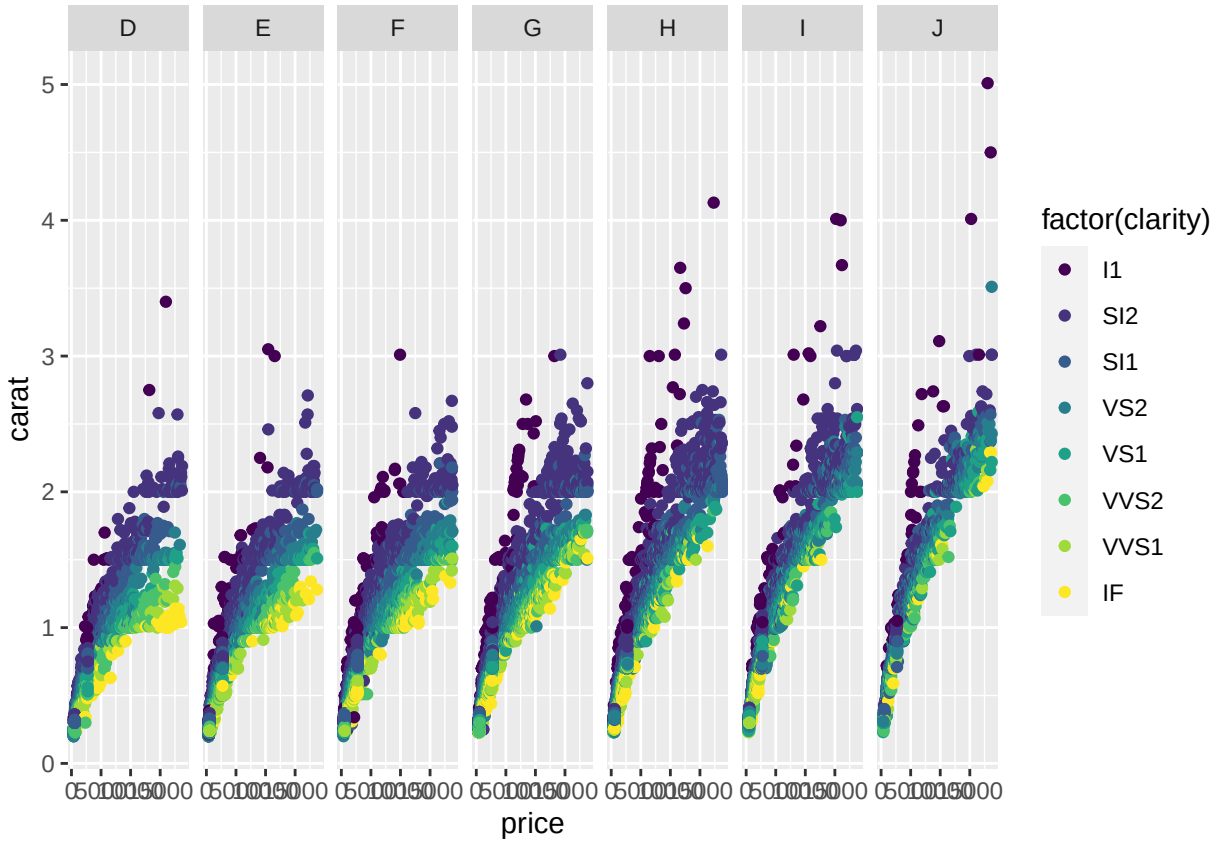
2022-09-13

Solutions

Question1

1a

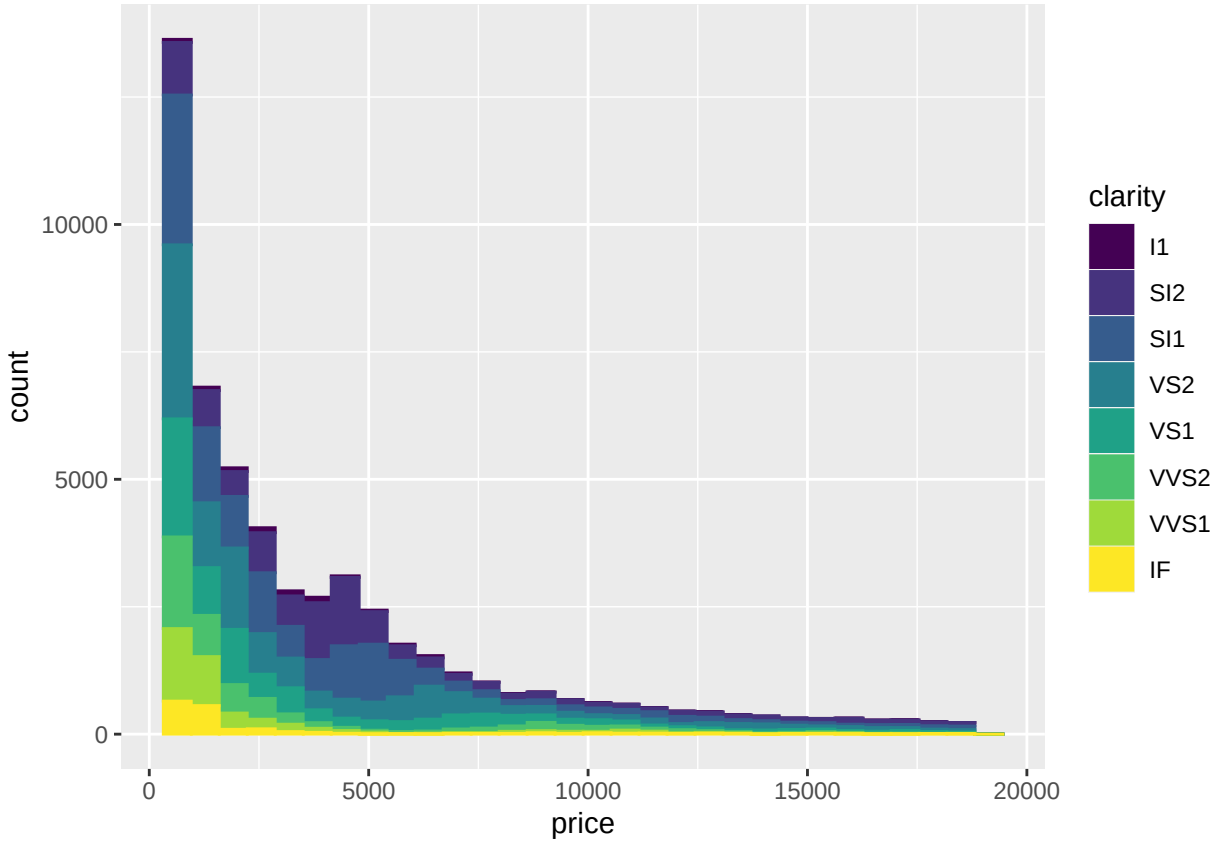
```
data(diamonds)
# head(diamonds)

ggplot(diamonds, aes(price, carat)) +
  geom_point(aes(color = factor(clarity))) +
  facet_grid(cols = vars(color))
```



1b

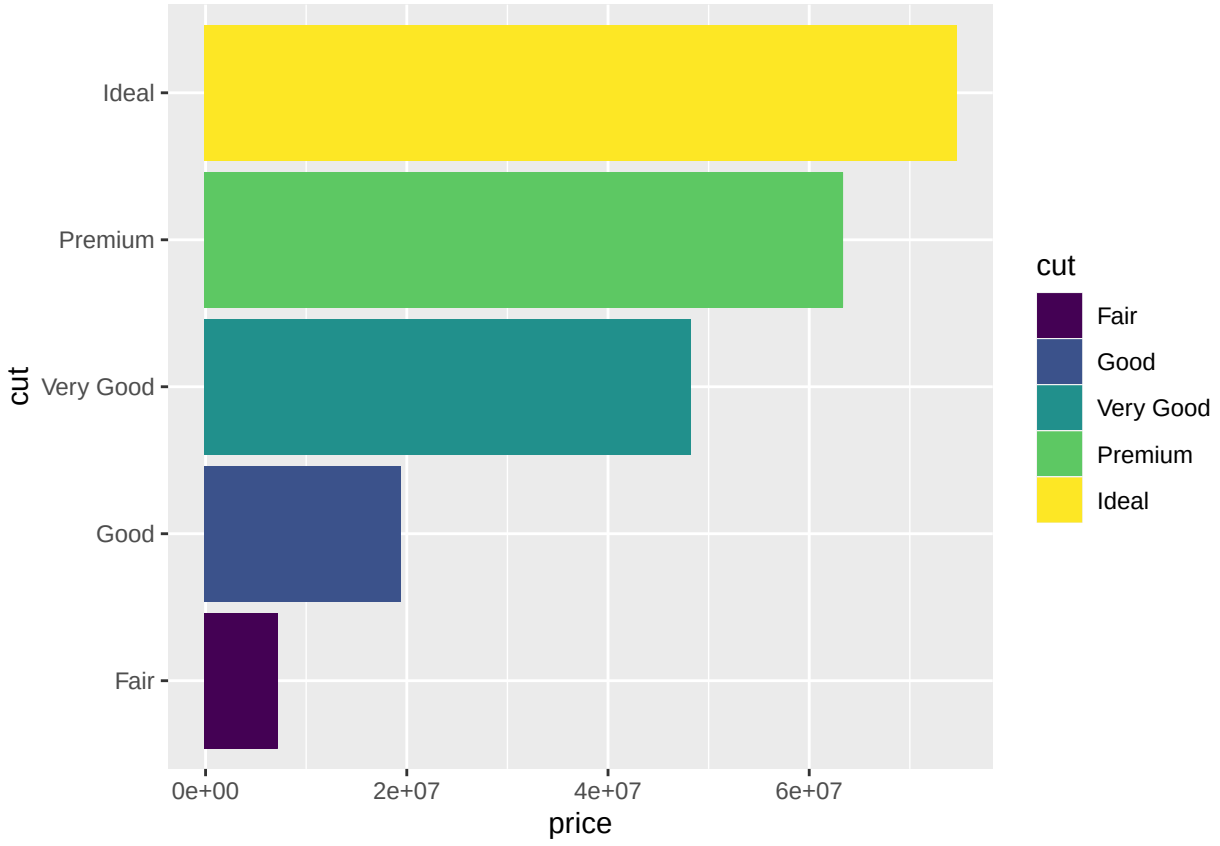
```
ggplot(diamonds, aes(price)) + geom_histogram(aes(color = clarity, fill = clarity))
```

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



1c

```
ggplot(diamonds, aes(price, cut)) + geom_bar(stat = "identity", aes(color = cut, fill = cut))
```



1d

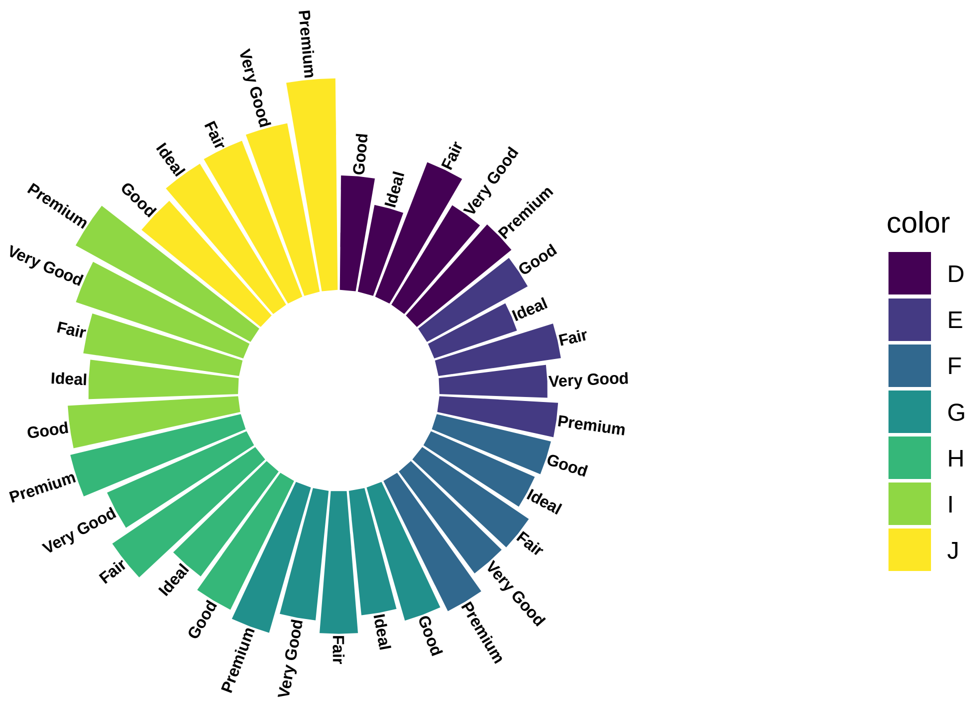
```
tempVec4 = c("Good", "Ideal", "Fair", "Very Good", "Premium")

dataframe11 = diamonds[order(match(diamonds$cut, tempVec4)),]
dataframe11$cut = factor(as.character(dataframe11$cut), levels = unique(dataframe11$cut))
dataframe12 <- dataframe11 %>%
  group_by(color, cut) %>%
  summarise(m = mean(price)) %>%
  ungroup() %>%
  mutate(id = 1:35)
```

```
## `summarise()` has grouped output by 'color'. You can override using the
## `.groups` argument.
```

```
label_data <- dataframe12
number_of_bar <- nrow(label_data)
angle <-  90 - 360 * (label_data$id-0.5) / number_of_bar

label_data$hjust<-ifelse( angle < -90, 1, 0)
label_data$angle<-ifelse(angle < -90, angle+180, angle)

ggplot(dataframe12, aes(x = as.factor(id), y = m, group = cut, fill = color)) +
```

```
geom_bar(stat = "identity", width = 0.88, position = position_dodge(0.99)) +
coord_polar(start = 0) + ylim(-3000, 9000)+ theme_minimal() +
theme(axis.text = element_blank(),axis.title = element_blank(),
      panel.grid = element_blank(),plot.margin = unit(rep(-1,4), "cm"))+
geom_text(data=label_data, aes(x=as.factor(id), y=m+40, label = cut, hjust = hjust),
          fontface="bold", size=2.2, size=0.5, angle= label_data$angle,inherit.aes = FALSE)
```

```
## Warning: Duplicated aesthetics after name standardisation: size
```

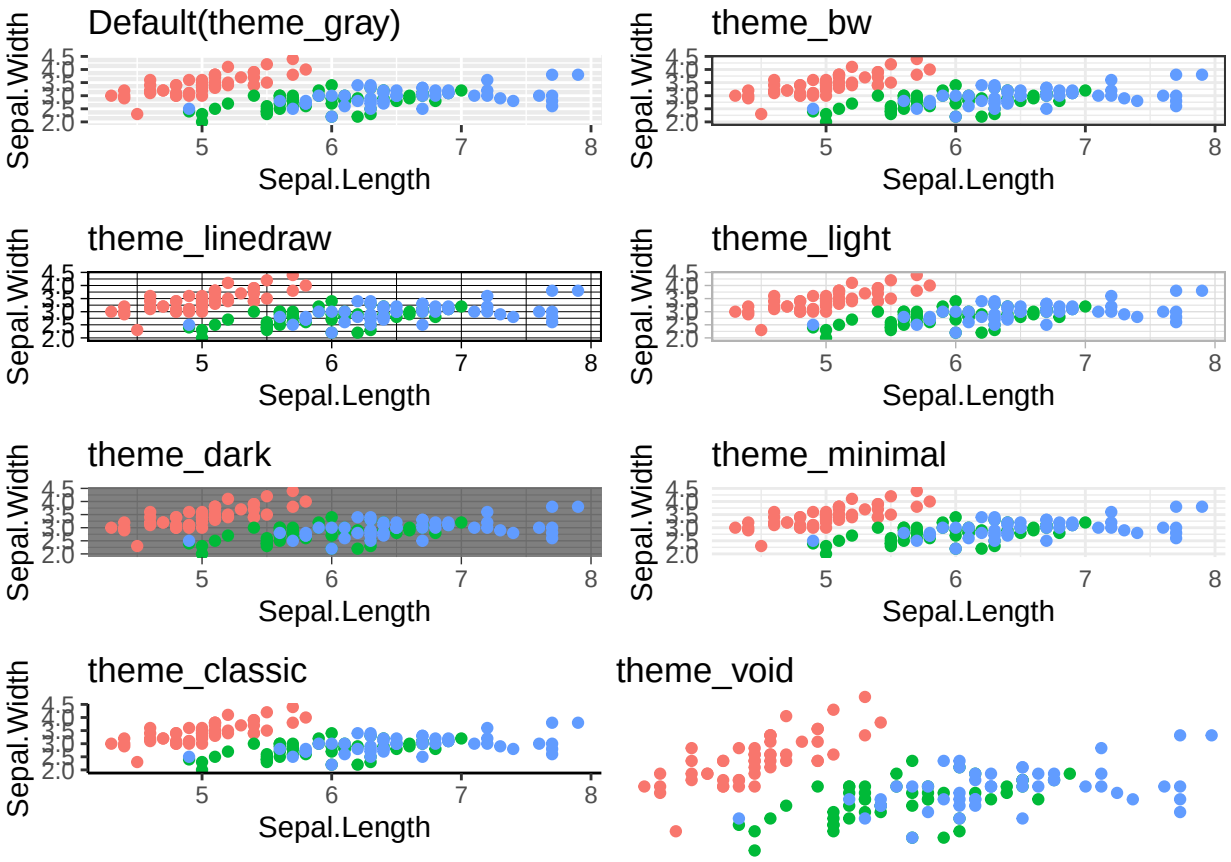


Question2

2a

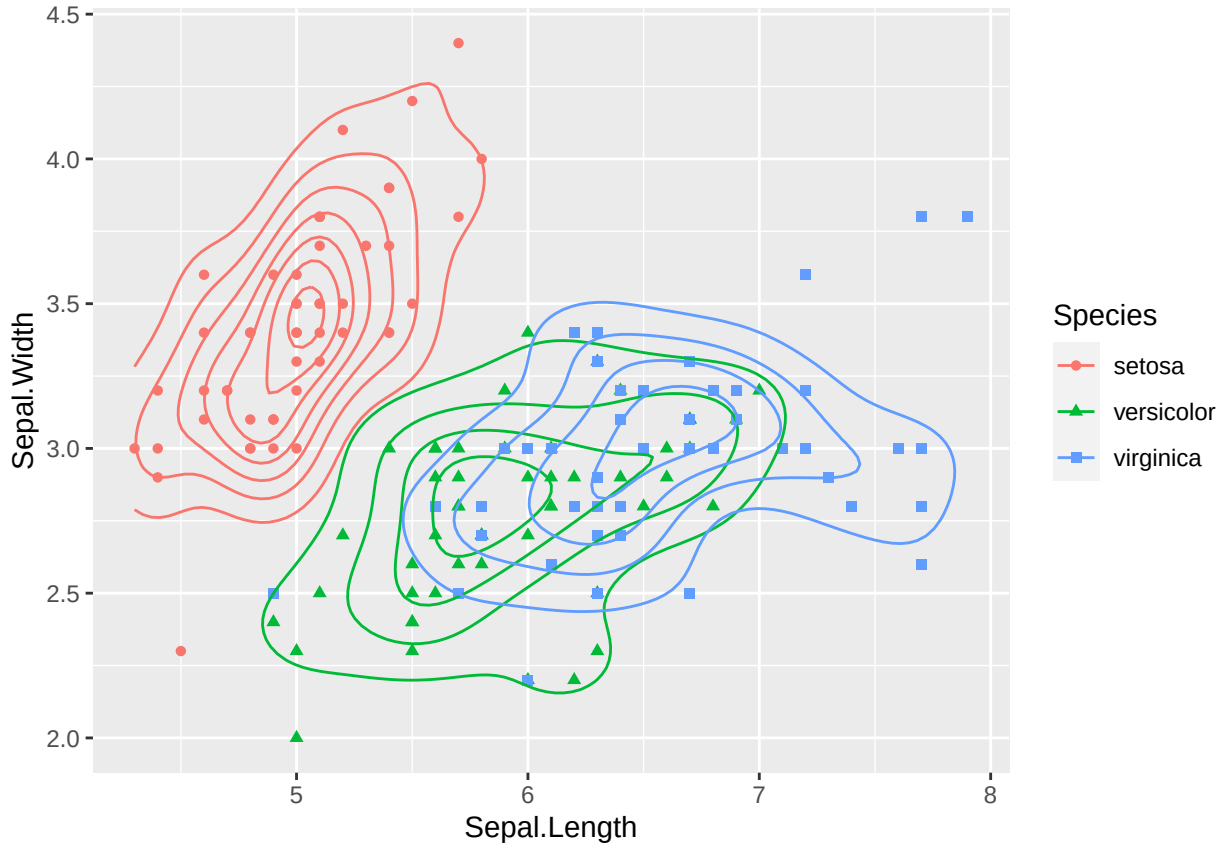
```
# install.packages("gridExtra")

data(iris)

# head(iris)
plot1 <- ggplot(iris, aes(Sepal.Length, Sepal.Width, color = Species)) +
labs(title = "Default(theme_gray)") + geom_point(show.legend = FALSE) + theme_gray()
```

```
plot2 <- ggplot(iris, aes(Sepal.Length, Sepal.Width, color = Species)) +
labs(title = "theme_bw") + geom_point(show.legend = FALSE) + theme_bw()

plot3 <- ggplot(iris, aes(Sepal.Length, Sepal.Width, color = Species)) +
labs(title = "theme_linedraw") + geom_point(show.legend = FALSE) + theme_linedraw()

plot4 <- ggplot(iris, aes(Sepal.Length, Sepal.Width, color = Species)) +
labs(title = "theme_light") + geom_point(show.legend = FALSE) + theme_light()

plot5 <- ggplot(iris, aes(Sepal.Length, Sepal.Width, color = Species)) +
labs(title = "theme_dark") + geom_point(show.legend = FALSE) + theme_dark()

plot6 <- ggplot(iris, aes(Sepal.Length, Sepal.Width, color = Species)) +
labs(title = "theme_minimal") + geom_point(show.legend = FALSE) + theme_minimal()

plot7 <- ggplot(iris, aes(Sepal.Length, Sepal.Width, color = Species)) +
labs(title = "theme_classic") + geom_point(show.legend = FALSE) + theme_classic()

plot8 <- ggplot(iris, aes(Sepal.Length, Sepal.Width, color = Species)) +
labs(title = "theme_void") + geom_point(show.legend = FALSE) + theme_void()


grid.arrange(plot1, plot2, plot3, plot4, plot5, plot6, plot7, plot8, ncol = 2 )
```



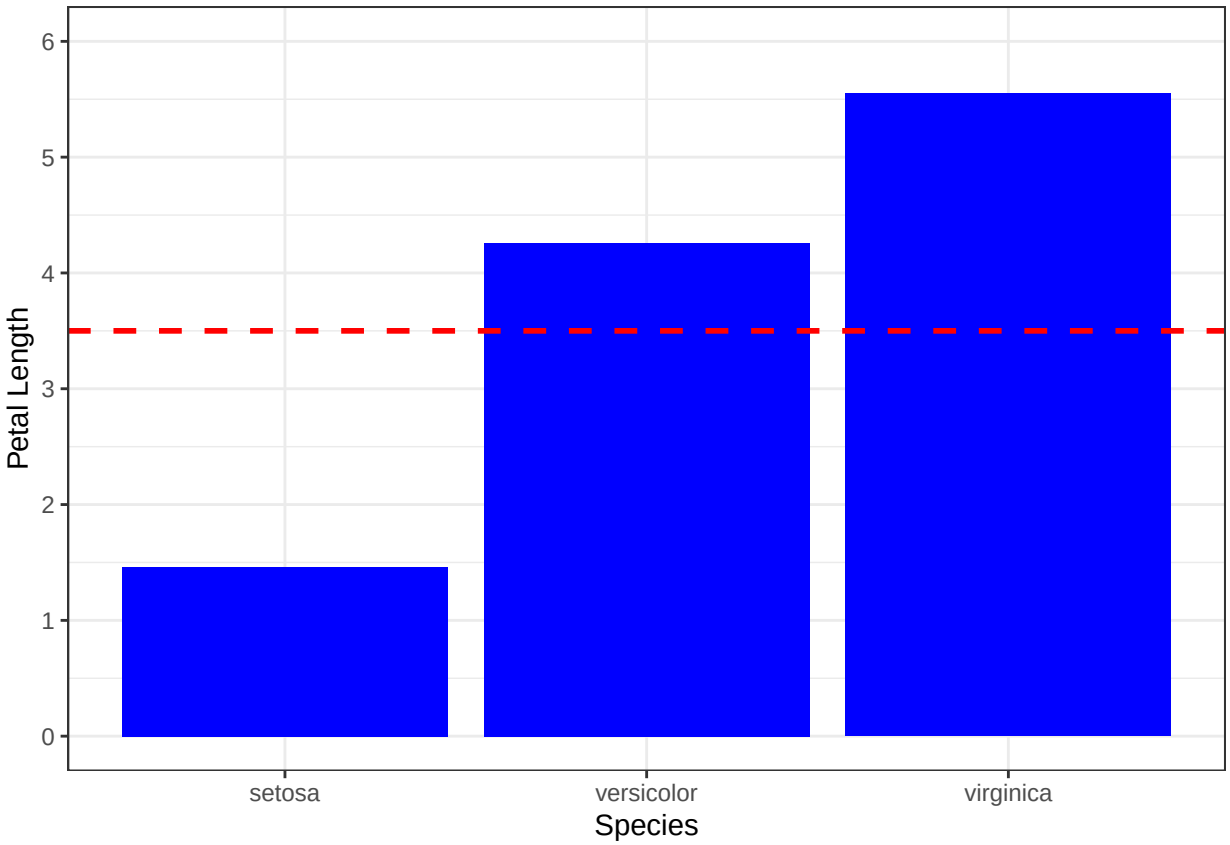
2b

```
ggplot(iris, aes(Sepal.Length, Sepal.Width, color = Species, shape = Species)) +
  geom_point() +
  geom_density2d()
```



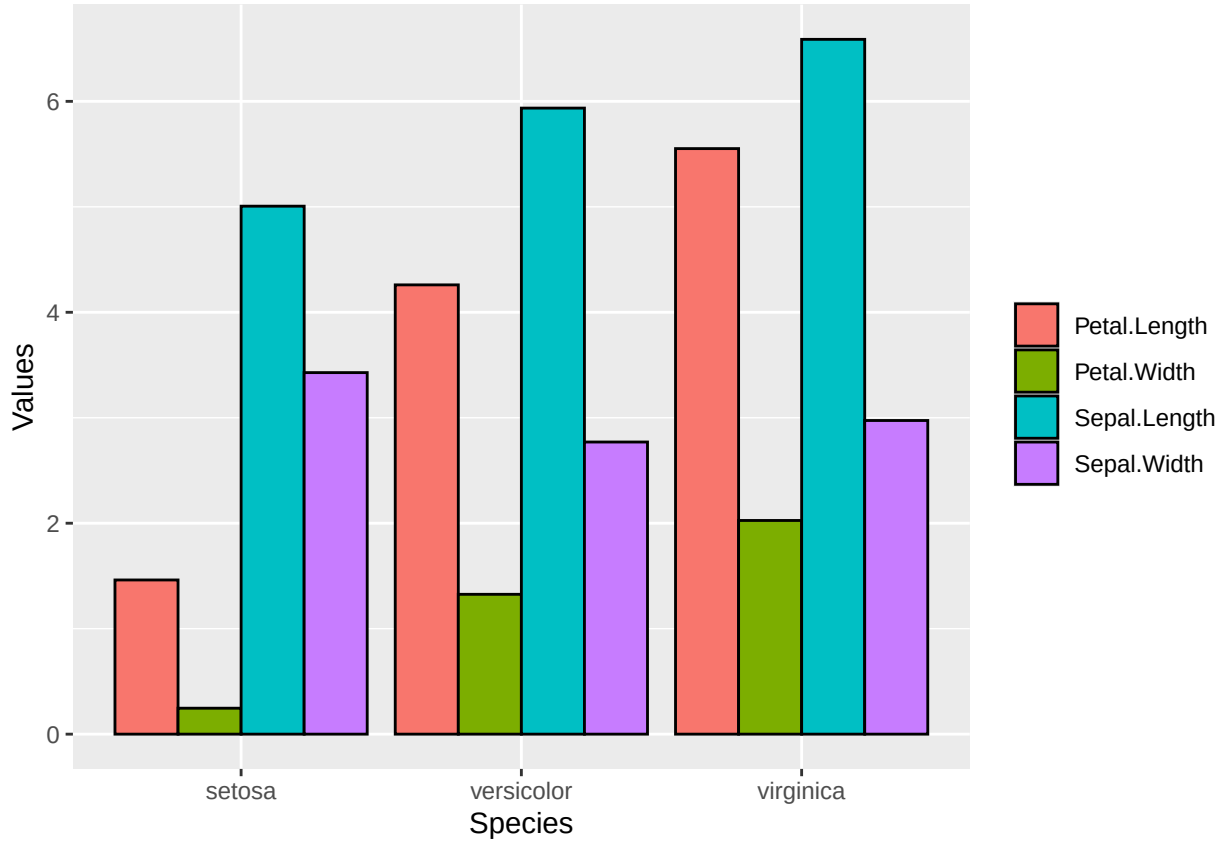
2c

```
dataframe1 = aggregate(iris[,1:4], by = list(iris$Species), FUN = mean)

ggplot(dataframe1, aes(x = Group.1, y = Petal.Length)) +
geom_bar(stat="identity",fill="blue") +
xlab('Species') + ylab('Petal Length') + theme_bw()+
scale_y_continuous(breaks = seq(0, 6, by=1), limits=c(0,6))+
geom_hline(yintercept=3.5, linetype="dashed", color = "red", size=1)
```



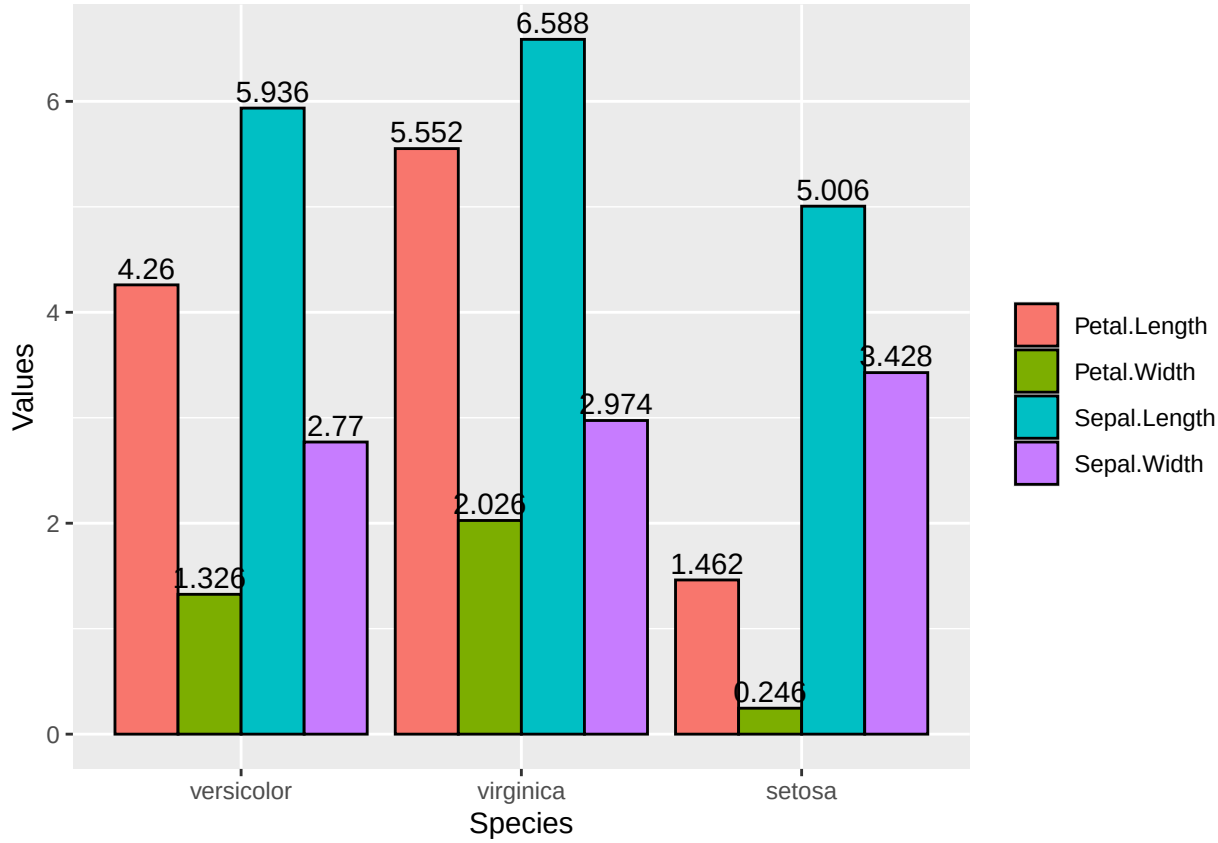
2d

```
dataframe2 <- melt(dataframe1, id.vars = c("Group.1"))
tempVec <- c("Petal.Length",
             "Petal.Width",
             "Sepal.Length",
             "Sepal.Width")

dataframe3 <- dataframe2[order(match(dataframe2$variable,
                                     tempVec)),]
dataframe3$variable <- factor(as.character(dataframe3$variable),
                              levels = unique(dataframe3$variable))

ggplot(dataframe3, aes(Group.1, value,
                       group = variable,
                       fill = variable)) +
  geom_bar(stat="identity",
           color = "black",
           position = "dodge") +
  xlab("Species") +
  ylab("Values") +
  scale_fill_discrete(name = NULL)
```



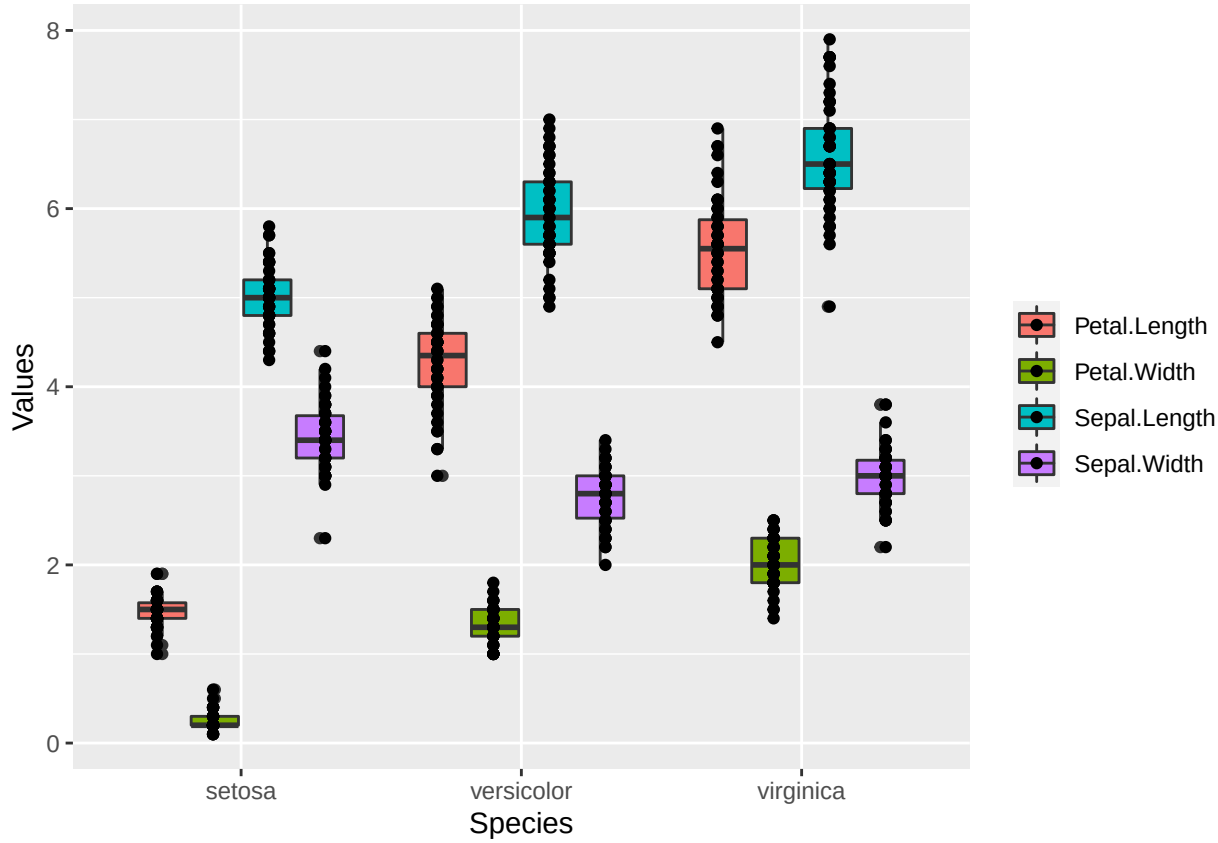
2e

```
tempVec2 <- c("versicolor","virginica","setosa")
dataframe4 <- dataframe3[order(match(dataframe3$Group.1, tempVec2)),]
dataframe4$Group.1 = factor(as.character(dataframe4$Group.1),
                            levels = unique(dataframe4$Group.1))

ggplot(dataframe4, aes(Group.1, value, group = variable,
                       fill = variable)) +
  geom_bar(stat="identity",color='black', position = "dodge") +
  xlab('Species') + ylab('Values') + scale_fill_discrete(name = NULL) +
  geom_text(aes(label = value), vjust = -0.3, size = 4,
            position = position_dodge(0.9))
```



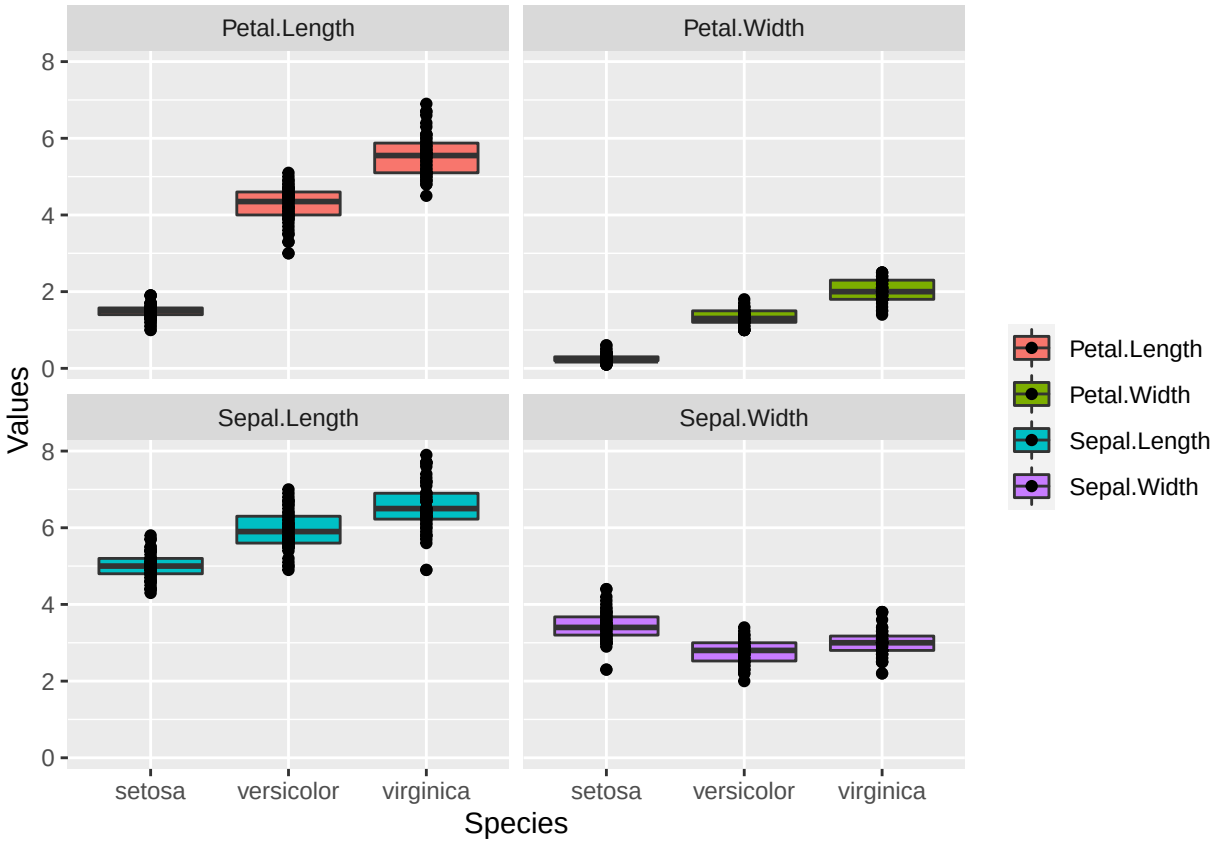
2f

```
dataframe5 <- melt(iris, id.vars = c("Species"))
tempVec <- c("Petal.Length","Petal.Width","Sepal.Length","Sepal.Width")

dataframe6 <- dataframe5[order(match(dataframe5$variable, tempVec)),]
dataframe6$variable <- factor(as.character(dataframe6$variable),
                              levels = unique(dataframe6$variable))

ggplot(dataframe6, aes(Species, value, fill = variable)) +
  geom_boxplot() + xlab("Species") + ylab("Values") +
  scale_fill_discrete(name = NULL) +
  geom_jitter(position = position_dodge(0.8))
```



2g

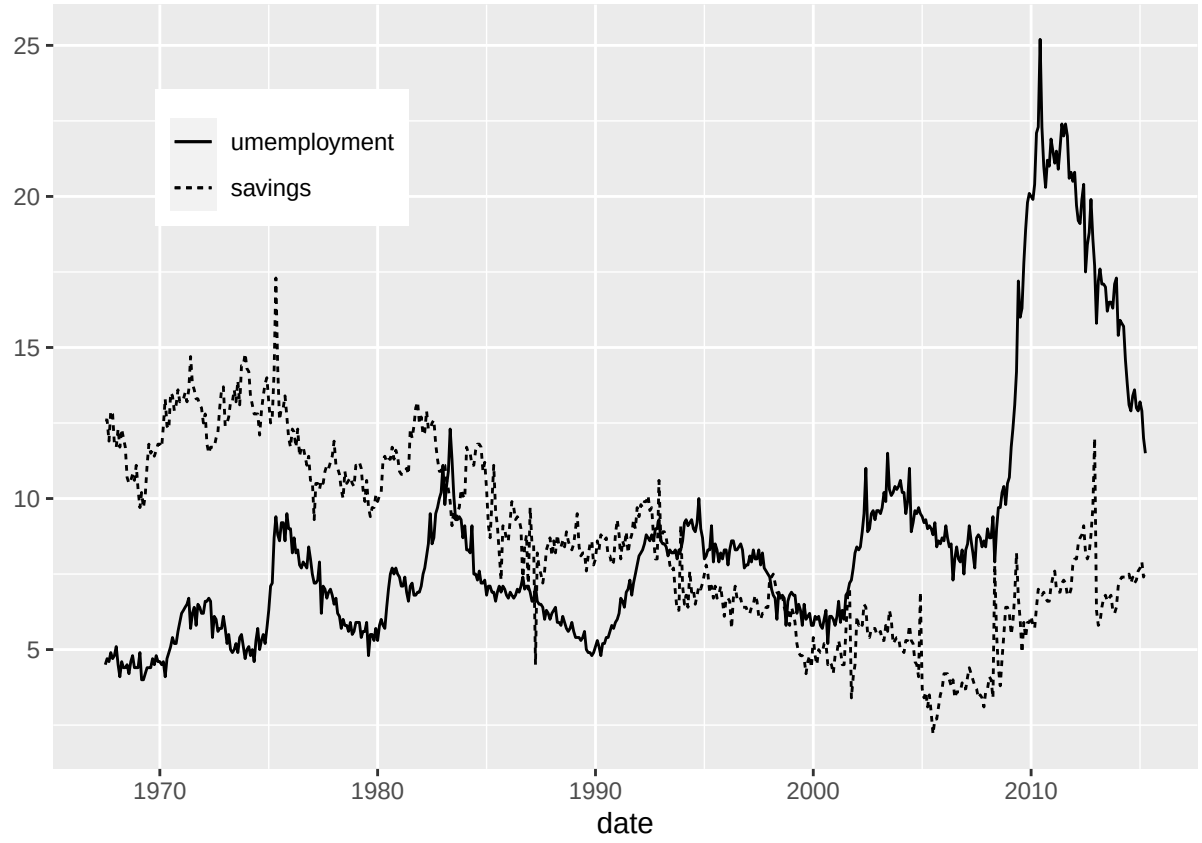
```
ggplot(dataframe6, aes(Species, value, fill = variable)) + geom_boxplot() +
  xlab('Species') + ylab('Values') +
  scale_fill_discrete(name=NULL) +
  geom_jitter(position=position_dodge(0.8)) +
  facet_wrap(~ variable, ncol = 2)
```



Question3

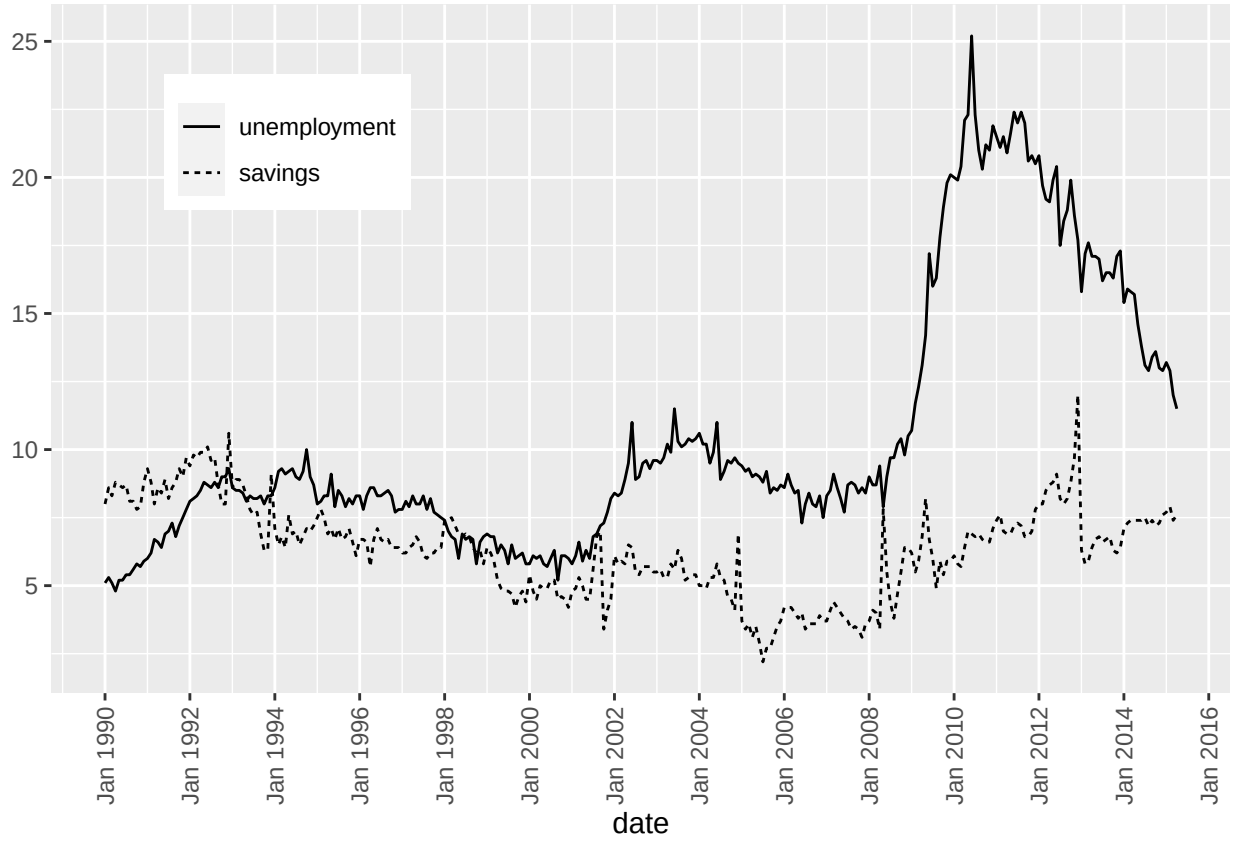
3a

```
data(economics)
# head(economics)

a <- ggplot(economics) + geom_line(aes(date, uempmed, linetype="dotted")) +
geom_line(aes(date, psavert, linetype="solid")) +
  theme(legend.position = c(0.2, 0.8)) +
ylab("")

b <- a + theme(legend.title = element_blank())
b + scale_linetype_discrete(breaks=c("dotted","solid"),
                            labels=c("umemployment","savings"))
```



3b

```
dataframe7 <- economics %>%
  dplyr::select(date, psavert, uempmed) %>%
  gather(key = "variable", value = "value", -date)
tempVec3 = c("uempmed", "psavert")
dataframe8 = dataframe7[order(match(dataframe7$variable, tempVec3)),]
dataframe8$variable = factor(as.character(dataframe8$variable),
                             levels = unique(dataframe8$variable))

ggplot(filter(dataframe8, date >= as.Date("1990-01-01")),
       aes(x = date, y = value)) + geom_line(aes(linetype = variable)) +
  ylab(NULL) + theme(legend.position = c(0.2, 0.8),
                     legend.title = element_blank()) +
  scale_linetype_discrete(labels = c("unemployment", "savings")) +
  scale_x_date(date_breaks = "2 years", date_labels = "%b %Y") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



Question4

4a

```
# install.packages("raster")
# install.packages("ncdf4")

dat <- nc_open("DMWC_G16.nc")
ws <- ncvar_get(dat, "wind_speed")
wd <- ncvar_get(dat, "wind_direction")
lat <- ncvar_get(dat, "lat")
lon <- ncvar_get(dat, "lon")
time <- ncvar_get(dat, "time")
press <- ncvar_get(dat, "pressure")
temp <- ncvar_get(dat, "temperature")
lza <- ncvar_get(dat, "local_zenith_angle")
sza <- ncvar_get(dat, "solar_zenith_angle")
DQF <- ncvar_get(dat, "DQF")
dataframe9 <- data.frame(ws,wd,lat,lon,time,press,temp,lza,sza,DQF)

head(dataframe9)
```

```
##   ws wd      lat       lon      time press temp      lza      sza DQF
```

```
## 1 NA NA 50.72388 -82.60785 656121674    NA   NA 58.55858 89.95837  13
## 2 NA NA 50.81353 -82.51603 656121674    NA   NA 58.60806 89.93052  13
## 3 NA NA 50.75671 -82.42097 656121674    NA   NA 58.56322 89.85365  13
## 4 NA NA 50.75293 -82.22757 656121674    NA   NA 58.52841 89.73582  13
## 5 NA NA 50.67696 -82.07603 656121674    NA   NA 58.44822 89.62068  13
## 6 NA NA 50.72203 -82.04516 656121674    NA   NA 58.47523 89.61632  13
```

4b

```
library(sf)
```

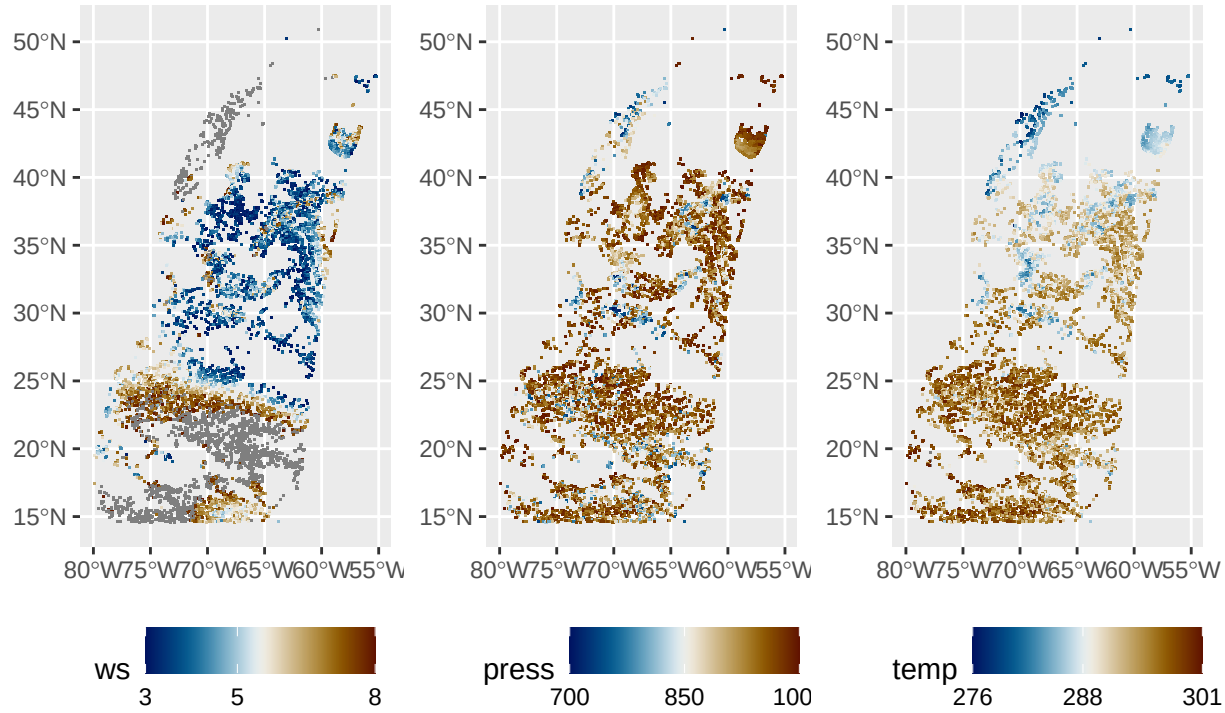
```
## Linking to GEOS 3.9.1, GDAL 3.4.3, PROJ 7.2.1; sf_use_s2() is TRUE
```

```
library(scico)
library(patchwork)
```

```
##
## Attaching package: 'patchwork'
```

```
## The following object is masked from 'package:raster':
##
##     area
```

```
df <- st_as_sf(dataframe9, coords = c("lon", "lat"), crs = 4326)
df <- subset(df, DQF==0)

plot9 <- ggplot(df, aes(color = ws)) + geom_sf(shape = 15, size = 0.1) +
  scale_color_scico(palette = "vik", breaks = c(3, 5, 8), limits = c(3,8)) +
  theme(legend.position =  'bottom', legend.direction = 'horizontal')

plot10 <- ggplot(df, aes(color = press)) + geom_sf(shape = 15, size = 0.1) +
  scale_color_scico(palette = "vik", breaks = c(700, 850,1000),
                    limits = c(700,1000)) +
  theme(legend.position =  'bottom', legend.direction = 'horizontal')

plot11 <- ggplot(df, aes(color = temp)) + geom_sf(shape = 15, size = 0.1) +
  scale_color_scico(palette = "vik", breaks = c(276, 288, 301),
                    limits = c(276,301)) +
  theme(legend.position =  'bottom', legend.direction = 'horizontal')


plot9 + plot10 + plot11
```



4c

```
library(tibble)

p <- tbl_df( df %>% pivot_longer(cols = c('ws','press','temp'),
                                 names_to = 'variable', values_to = 'value'))
```

```
## Warning: `tbl_df()` was deprecated in dplyr 1.0.0.
## Please use `tibble::as_tibble()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was generated.
```

```
head(p)
```

```
## # A tibble: 6 x 8
##      wd       time   lza   sza   DQF            geometry variable  value
##   <dbl>      <dbl> <dbl> <dbl> <int>          <POINT [°]> <chr>     <dbl>
## 1  209. 656121674.  60.0  77.1     0 (-60.31003 50.88714) ws        29.6
## 2  209. 656121674.  60.0  77.1     0 (-60.31003 50.88714) press    746.
## 3  209. 656121674.  60.0  77.1     0 (-60.31003 50.88714) temp     280.
## 4  263. 656121674.  58.7  78.3     0  (-63.0746 50.24621) ws         3.42
## 5  263. 656121674.  58.7  78.3     0  (-63.0746 50.24621) press    989.
## 6  263. 656121674.  58.7  78.3     0  (-63.0746 50.24621) temp     278.
```

4d

```
df$var1 <- "ws"
df$var2 <- "press"
df$var3 <- "temp"

plot12 <- ggplot(df, aes(color = ws)) + geom_sf(shape = 15, size = 0.1) +
  scale_color_scico(palette = "vik", breaks = c(10,20)) +
  theme(plot.title=element_text(hjust = 0.5, ), legend.position =  'bottom',
        legend.direction = 'horizontal', legend.title = element_blank()) +
  facet_grid(. ~ var1)

plot13 <- ggplot(df, aes(color = press)) + geom_sf(shape = 15, size = 0.1) +
  scale_color_scico(palette = "vik", breaks = c(800, 900))  +
  theme(legend.position =  'bottom', legend.direction = 'horizontal',
        legend.title = element_blank()) + facet_grid(. ~ var2)

plot14 <- ggplot(df, aes(color = temp)) + geom_sf(shape = 15, size = 0.1) +
  scale_color_scico(palette = "vik", breaks = c(280, 290, 300),
                    limits = c(280,300)) +
  theme(legend.position =  'bottom', legend.direction = 'horizontal',
        legend.title = element_blank())+ facet_grid(. ~ var3)


plot13+ plot14+ plot12
```