260Project_Jingyi

Jingyi Huang

2018/12/2

Exploratory Analysis

```
library(corrplot)
```

```
## corrplot 0.84 loaded
```

```
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(tidyverse)
```

```
## -- Attaching packages ---------------------------------------- tidyverse 1.2.1 --
```

```
## v ggplot2 3.1.0     v readr   1.1.1
## v tibble  1.4.2     v purrr   0.2.5
## v tidyr   0.8.1     v stringr 1.3.1
## v ggplot2 3.1.0     v forcats 0.3.0
```

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

```
library(ggjoy)
```

```
## Loading required package: ggridges
```

```
##
## Attaching package: 'ggridges'
```

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

```
## The ggjoy package has been deprecated. Please switch over to the
## ggridges package, which provides the same functionality. Porting
## guidelines can be found here:
## https://github.com/clauswilke/ggjoy/blob/master/README.md
```

```
library(devtools)
library(ggplot2)
```

```
#food <- read.csv("data/food_coded_clean.csv",stringsAsFactors = FALSE)
#food_numeric <- food %>% select_if(is.numeric)
```

```
#food_numeric <- food_numeric[complete.cases(food_numeric), ]

food=breakfast_clean <- readRDS("~/Desktop/BST260/data/breakfast_clean.rds")
class(food$comfort_food_reasons_coded)
```

```
## [1] "factor"
```

```
class(food$GPA)
```

```
## [1] "numeric"
```

```
class(food$weight)
```

```
## [1] "factor"
```

need to delete comfort food reasons.1 (pretty much identical to comfort food reasons)

```
#f3=food%>%
  #select(as.numeric(comfort_food_reasons_coded),as.numeric(calories_scone))

#F3 <- cor(f3)
#corrplot(F3, method = "number")
```

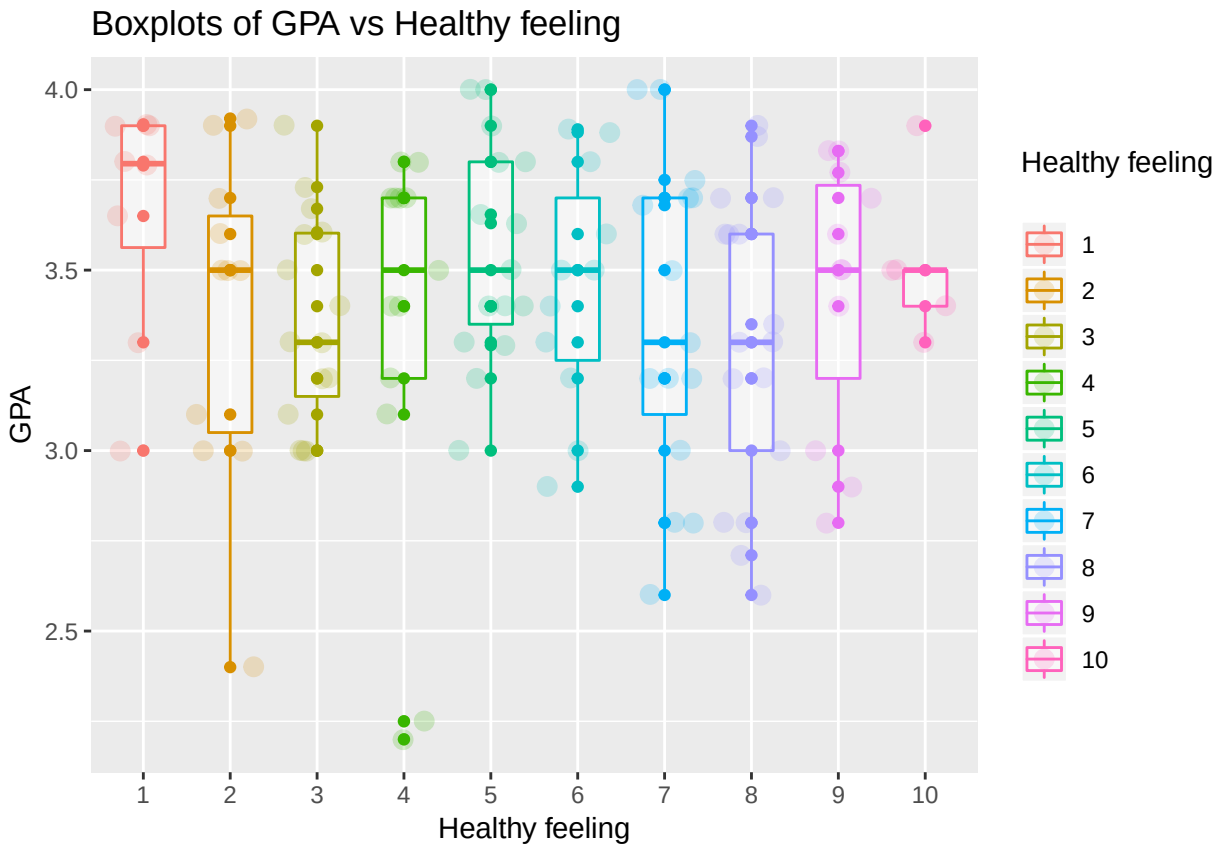
Health feeling =1 marks the people who are most confident that they are healthy. As we can see from the boxplots, people who are the most confident about their health condition associate with higher GPA (the median is the highest among the 10 rating categories).

```
food %>%
  select(healthy_feeling,GPA)%>%
    ggplot(aes( x=as.factor(healthy_feeling),y=GPA, color=as.factor(healthy_feeling))) +
    geom_boxplot(width=0.5, alpha=0.5) +
  geom_jitter(size=3,alpha=.2)+
  #geom_violin(width=1.4)+
  labs(color = "Healthy feeling\n")+
   geom_point(show.legend = FALSE) +
    xlab("Healthy feeling") +
    ylab("GPA") +
    ggtitle("Boxplots of GPA vs Healthy feeling")
```

```
## Warning: Removed 4 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

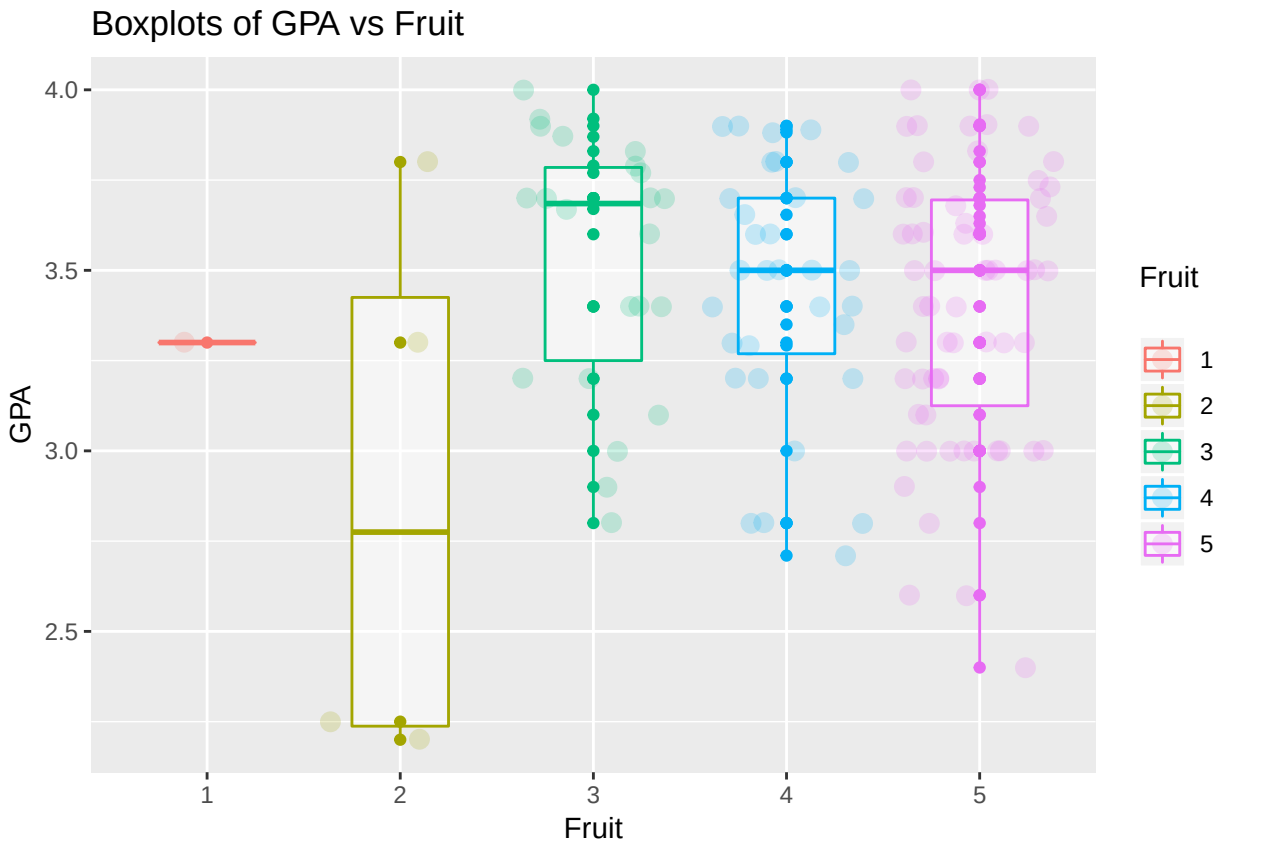



```
food %>%
  select(fruit_day,GPA)%>%
    ggplot(aes( x=as.factor(fruit_day),y=GPA, color=as.factor(fruit_day))) +
    geom_boxplot(width=0.5, alpha=0.5) +
  geom_jitter(size=3,alpha=.2)+
  #geom_violin(width=1.4)+
  labs(color = "Fruit\n")+
 geom_point(show.legend = FALSE) +
    xlab("Fruit") +
    ylab("GPA") +
    ggtitle("Boxplots of GPA vs Fruit")
```

```
## Warning: Removed 4 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

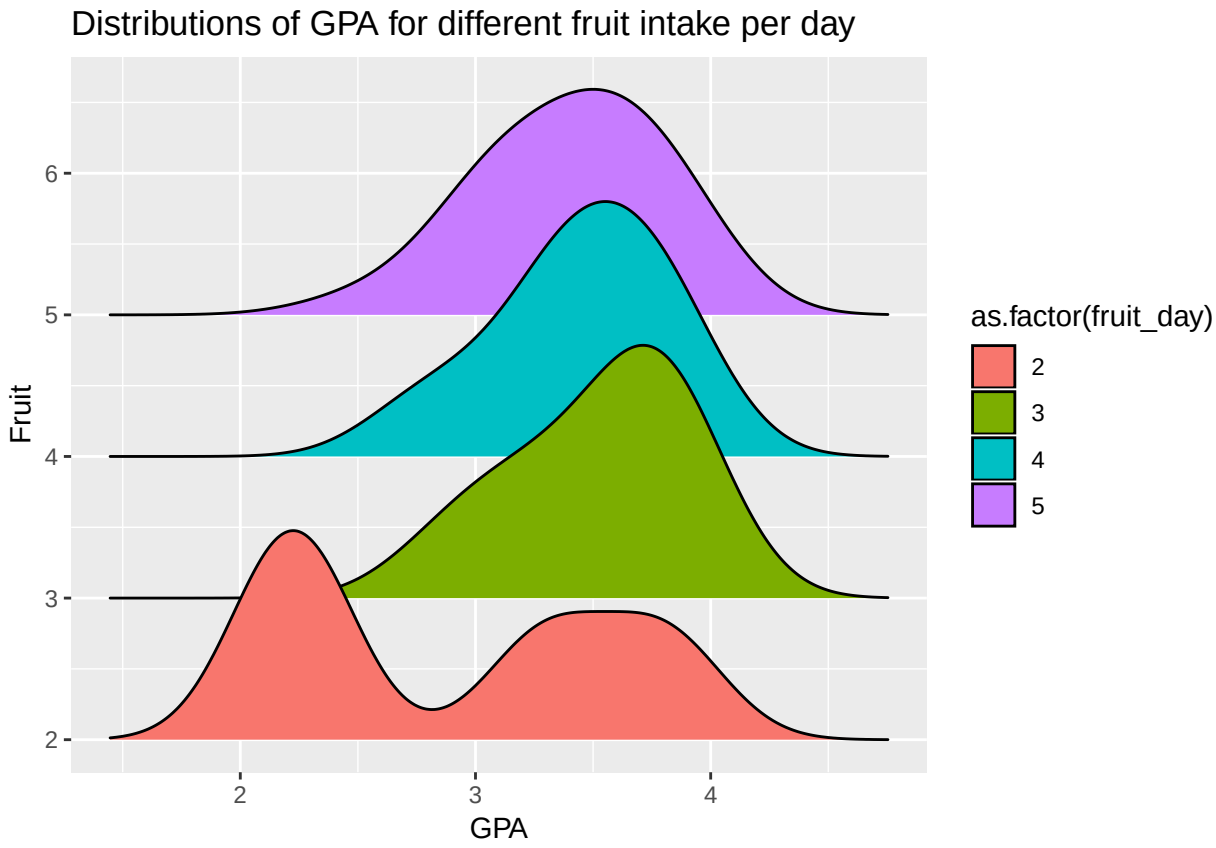
```
## Warning: Removed 4 rows containing missing values (geom_point).
```




```
food %>%
  select(fruit_day,GPA)%>%
    ggplot(aes( y=fruit_day,x=GPA, group=as.factor(fruit_day),fill=as.factor(fruit_day)))+
  geom_density_ridges()+
  labs(color = "Fruit\n")+
    ylab("Fruit") +
    ggtitle("Distributions of GPA for different fruit intake per day")
```

```
## Picking joint bandwidth of 0.251
```

```
## Warning: Removed 4 rows containing non-finite values (stat_density_ridges).
```



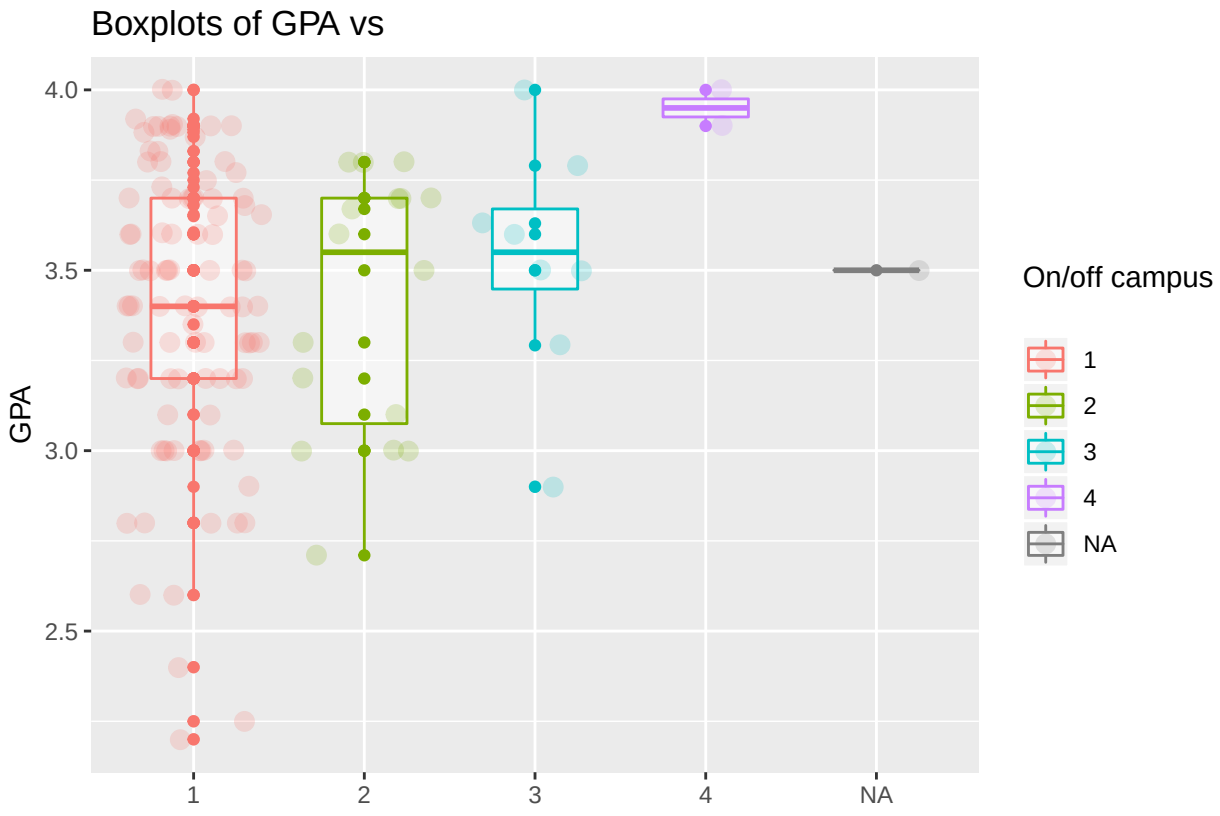
on/off campus

```
food %>%
  select(on_off_campus,GPA)%>%
    ggplot(aes( x=as.factor(on_off_campus),y=GPA, color=as.factor(on_off_campus))) +
    geom_boxplot(width=0.5, alpha=0.5) +
  geom_jitter(size=3,alpha=.2)+
  #geom_violin(width=1.4)+
  labs(color = "On/off campus\n")+
 geom_point(show.legend = FALSE) +
    xlab("") +
    ylab("GPA") +
    ggtitle("Boxplots of GPA vs ")
```

```
## Warning: Removed 4 rows containing non-finite values (stat_boxplot).
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```

```
## Warning: Removed 4 rows containing missing values (geom_point).
```


Boxplots of GPA vs
GPA
4.0
3.5
3.0
2.5
1
2
3
4
NA
On/off campus
1
2
3
4
NA