有用なR分析技術

③最高の視覚化Rパッケージ “ggplot2” 2

練習に使うデータの紹介

生徒の肥満度と運動時間（分）を示しているデータセット

A tibble: 20 × 3

	`Student Number`	`Obesity index`	Workout
	<chr>	<chr>	<dbl>
1	S_001	Obesity	60
2	S_002	Obesity	30
3	S_003	Obesity	120
4	S_004	Obesity	60
5	S_005	Obesity	0
6	S_006	Obesity	60
7	S_007	Obesity	120
8	S_008	Obesity	180
9	S_009	Obesity	60
10	S_010	Obesity	45
11	S_011	Normal	30
12	S_012	Normal	105
13	S_013	Normal	90
14	S_014	Normal	135
15	S_015	Normal	120
16	S_016	Normal	90
17	S_017	Normal	85
18	S_018	Normal	100
19	S_019	Normal	120
20	S_020	Normal	60

肥満と正常の二つのグループに分けて連続数値である運動時間をy軸とする箱ひげ図を作成する

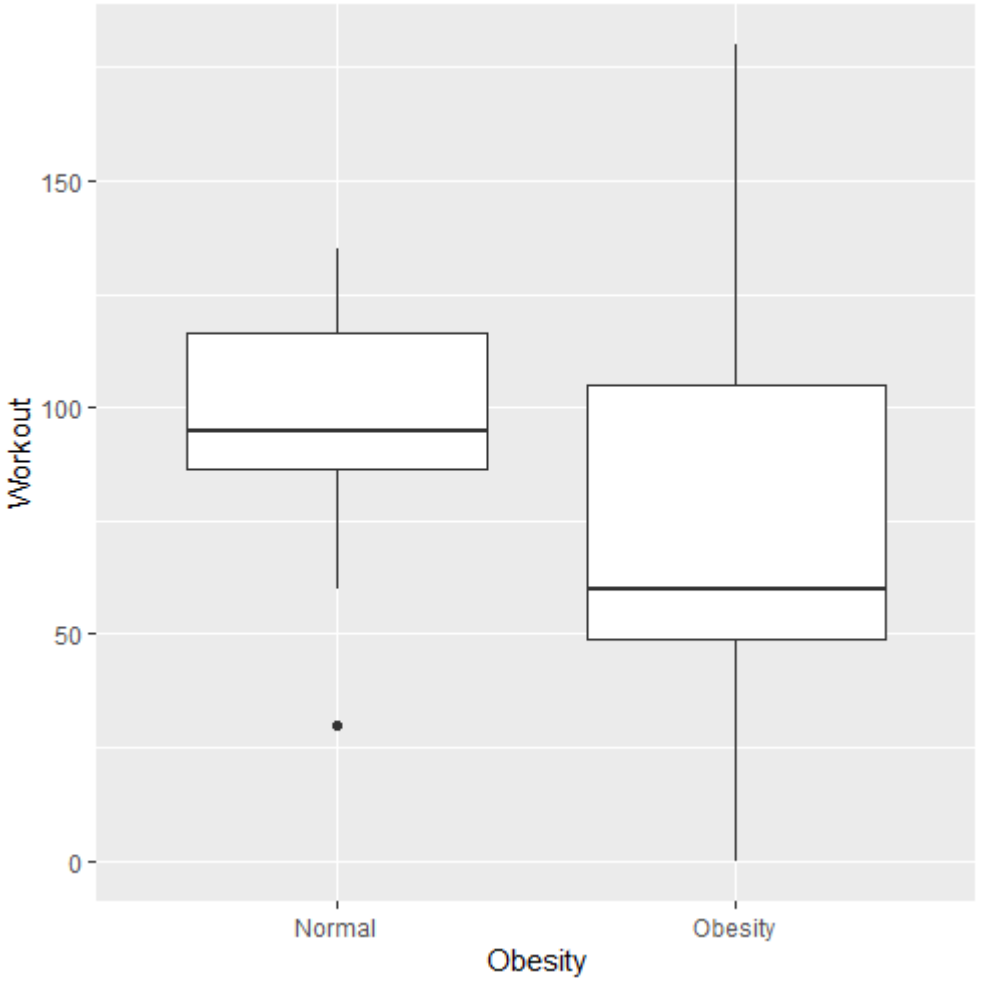
データの全処理

```r
library(tidyverse)
library(nparcomp)
library(readxl)

getwd()

## Load dataset ##

obesity_data ← readxl::read_excel("./03.Example data (最高の視覚化Rパッケージggplot2_2).xlsx")
View(obesity_data)
obesity_data


## Rename columns ##

obesity_data_renamed ← obesity_data %>% dplyr::rename(Student = `Student Number`,
                                                      Obesity = `Obesity index`)

obesity_data_renamed


## Preprocessing for graph ##

obesity_data_renamed$Obesity ← factor(obesity_data_renamed$Obesity, levels=c("Normal", "Obesity"))
obesity_data_renamed
```

NormalとObesityをfactorに変更

グラフの基本型を作る

```
ggplot(obesity_data_renamed, aes(x = Obesity, y = Workout)) +
  geom_boxplot()
```

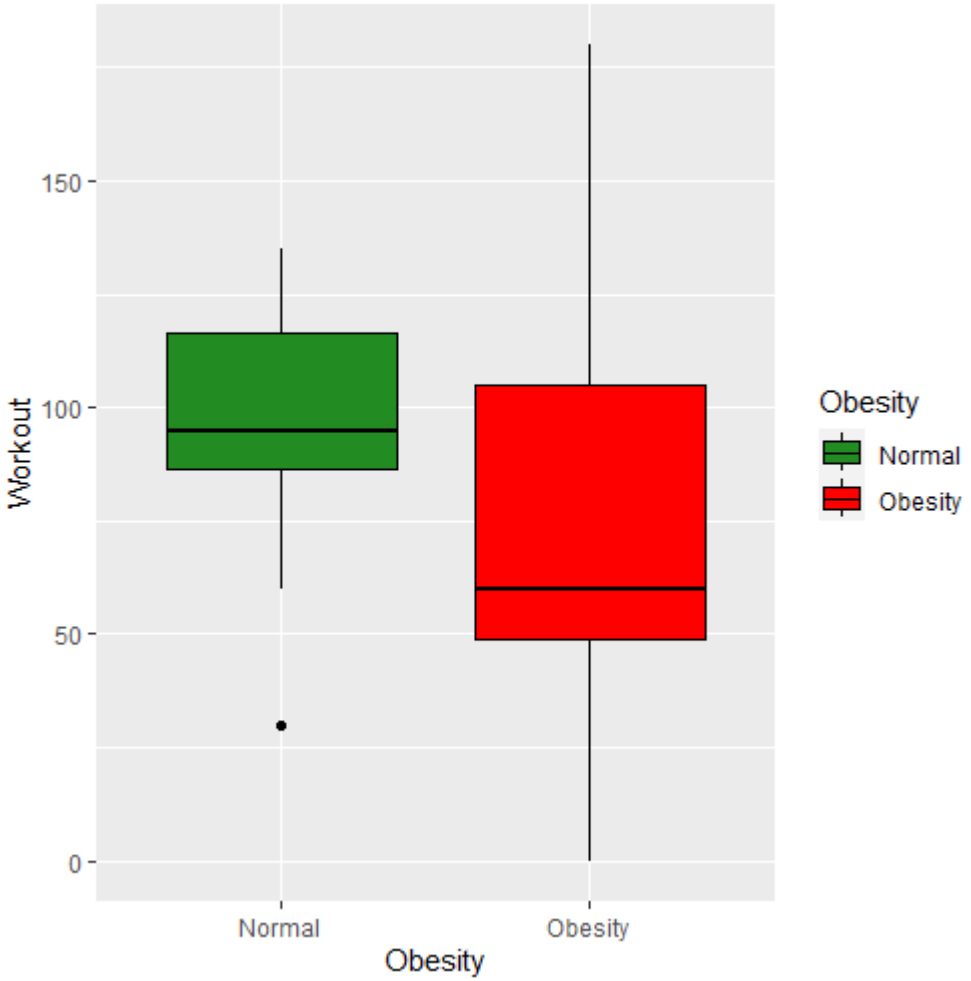


グラフセッティング

①色を設定する

```
ggplot(obesity_data_renamed, aes(x = Obesity, y = Workout, fill = Obesity)) +
  geom_boxplot(color = 'black', outlier.shape = NA) +
  scale_fill_manual(values = c('forestgreen', 'red'))
```

Normalは緑色、Obesityは赤色に設定

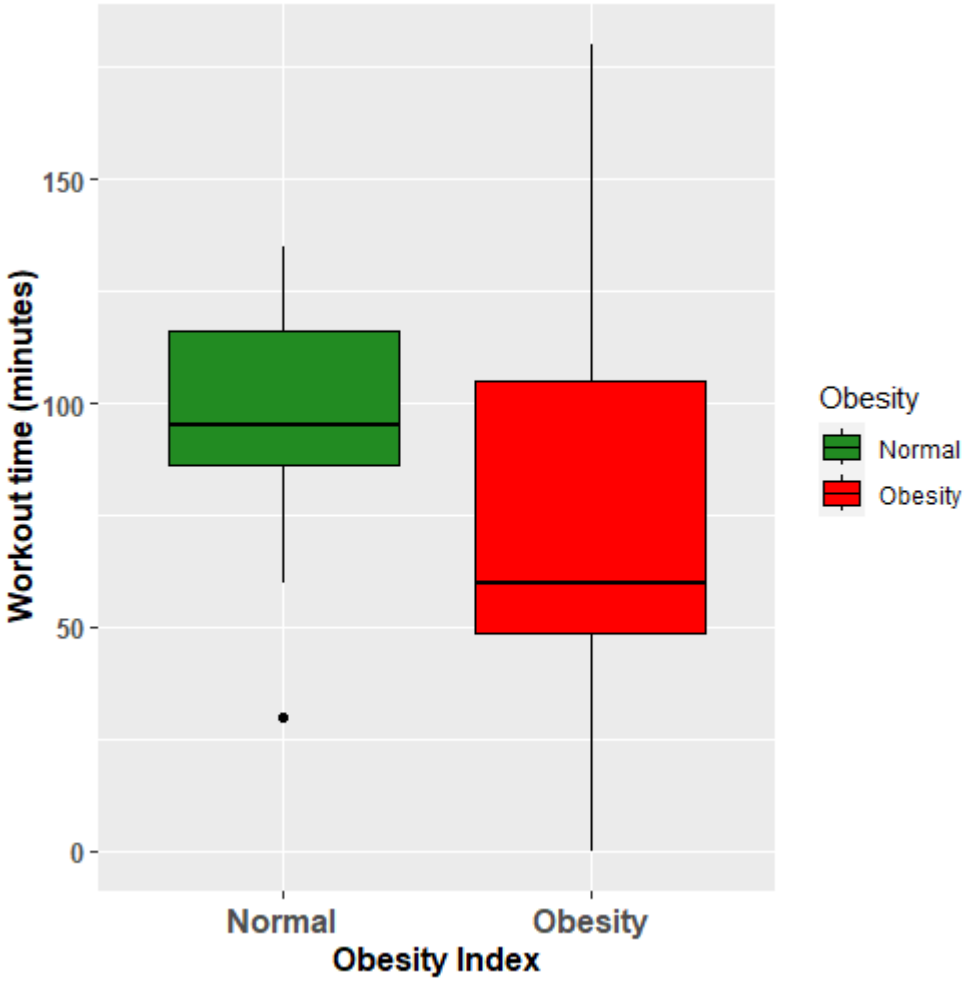


グラフセッティング

②軸名と目盛のデザインを変更する

```r
# Change axes names and tick settings

ggplot(obesity_data_renamed, aes(x = Obesity, y = Workout, fill = Obesity)) +
  geom_boxplot(color = 'black', outlier.shape = NA) +
  scale_fill_manual(values = c('forestgreen', 'red')) +
  xlab('Obesity Index') +
  ylab('Workout time (minutes)') +
  theme(axis.title.x = element_text(size = 12, face = 'bold'),
        axis.title.y = element_text(size = 12, face = 'bold'),
        axis.text.x = element_text(size = 12, face = 'bold'),
        axis.text.y = element_text(size = 10, face = 'bold'))
```

軸名と目盛のサイズの変更及びボルド体に



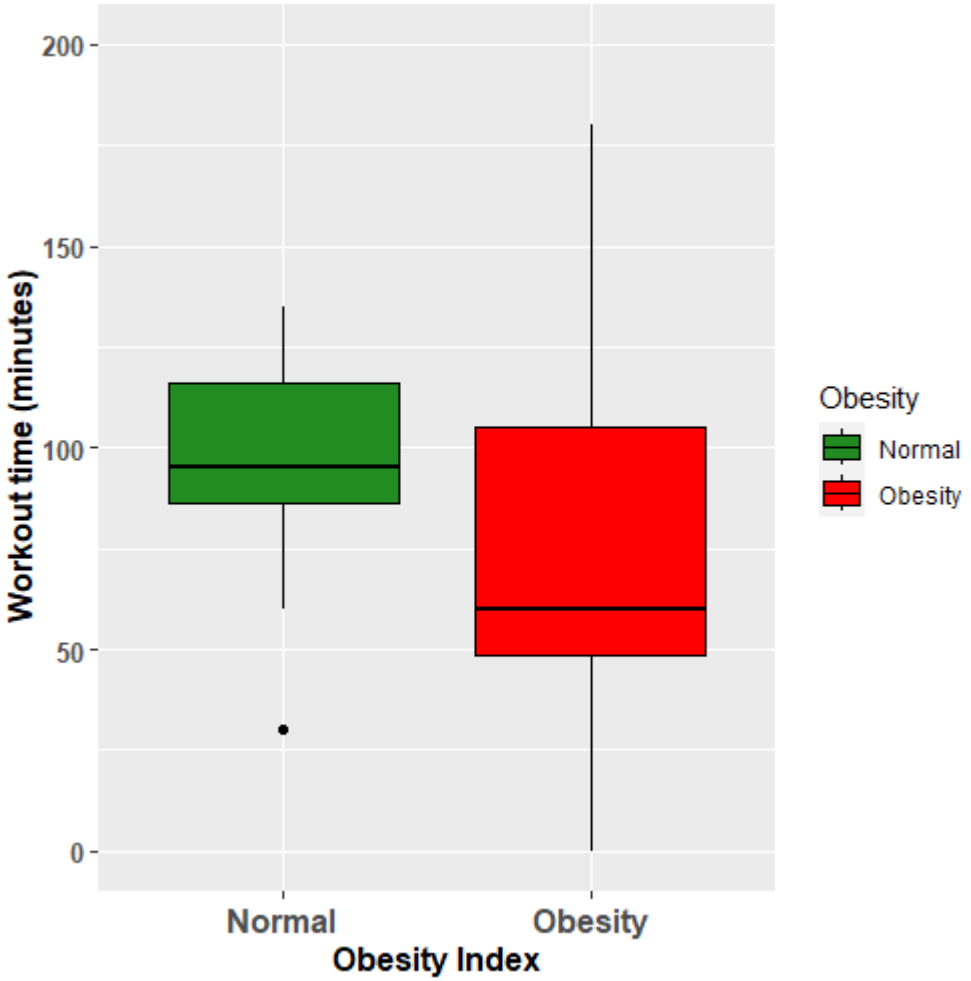
グラフセッティング

③Y軸の範囲を設定

Y軸の最大値を確認するため

```
max(obesity_data_renamed$Workout) # 180

ggplot(obesity_data_renamed, aes(x = Obesity, y = Workout, fill = Obesity)) +
  geom_boxplot(color = 'black', outlier.shape = NA) +
  scale_fill_manual(values = c('forestgreen', 'red')) +
  xlab('Obesity Index') +
  ylab('Workout time (minutes)') +
  theme(axis.title.x = element_text(size = 12, face = 'bold'),
        axis.title.y = element_text(size = 12, face = 'bold'),
        axis.text.x = element_text(size = 12, face = 'bold'),
        axis.text.y = element_text(size = 10, face = 'bold')) +
  scale_y_continuous(limits = c(0,200))
```

Y軸の範囲を0から200までに設定

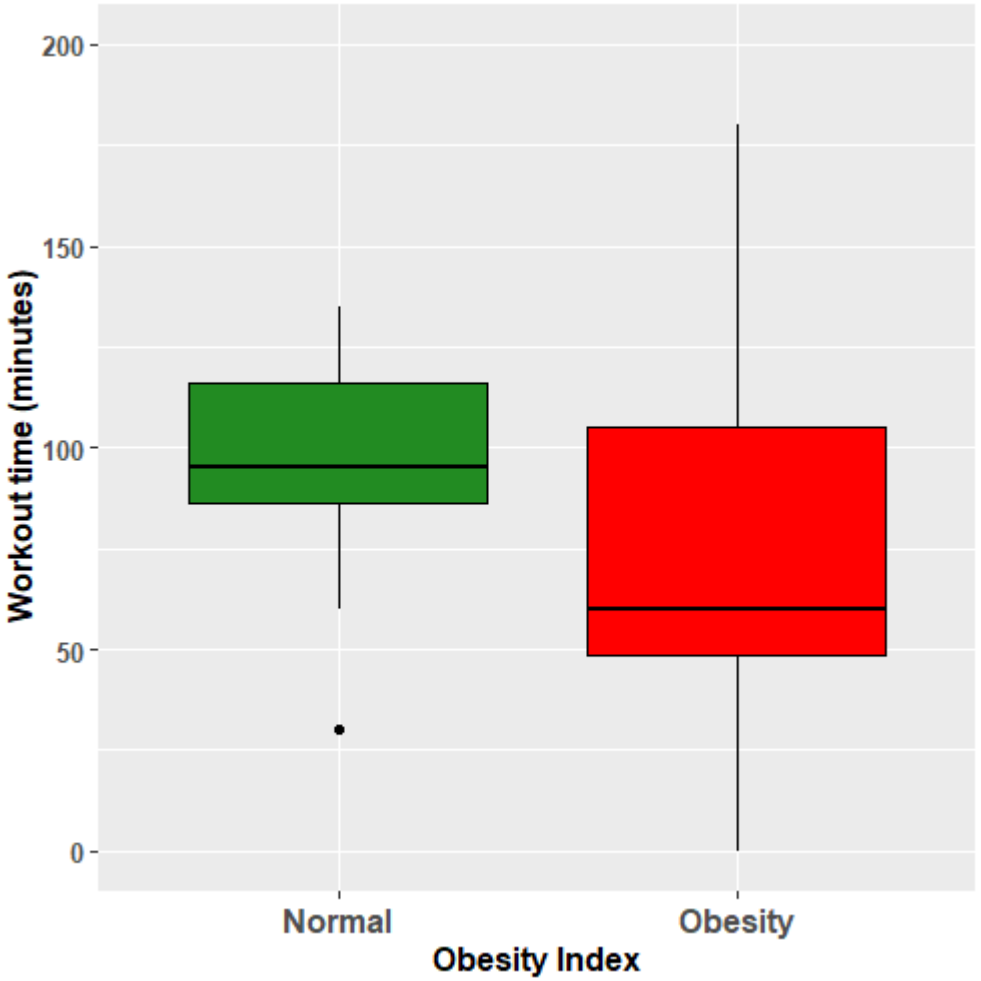


グラフセッティング

④凡例を削除する

```
ggplot(obesity_data_renamed, aes(x = Obesity, y = Workout, fill = Obesity)) +
  geom_boxplot(color = 'black', outlier.shape = NA) +
  scale_fill_manual(values = c('forestgreen', 'red')) +
  xlab('Obesity Index') +
  ylab('Workout time (minutes)') +
  theme(axis.title.x = element_text(size = 12, face = 'bold'),
        axis.title.y = element_text(size = 12, face = 'bold'),
        axis.text.x = element_text(size = 12, face = 'bold'),
        axis.text.y = element_text(size = 10, face = 'bold')) +
  scale_y_continuous(limits = c(0,200)) +
  theme(legend.position = "none")
```

このグラフでは色は重要ではないため削除

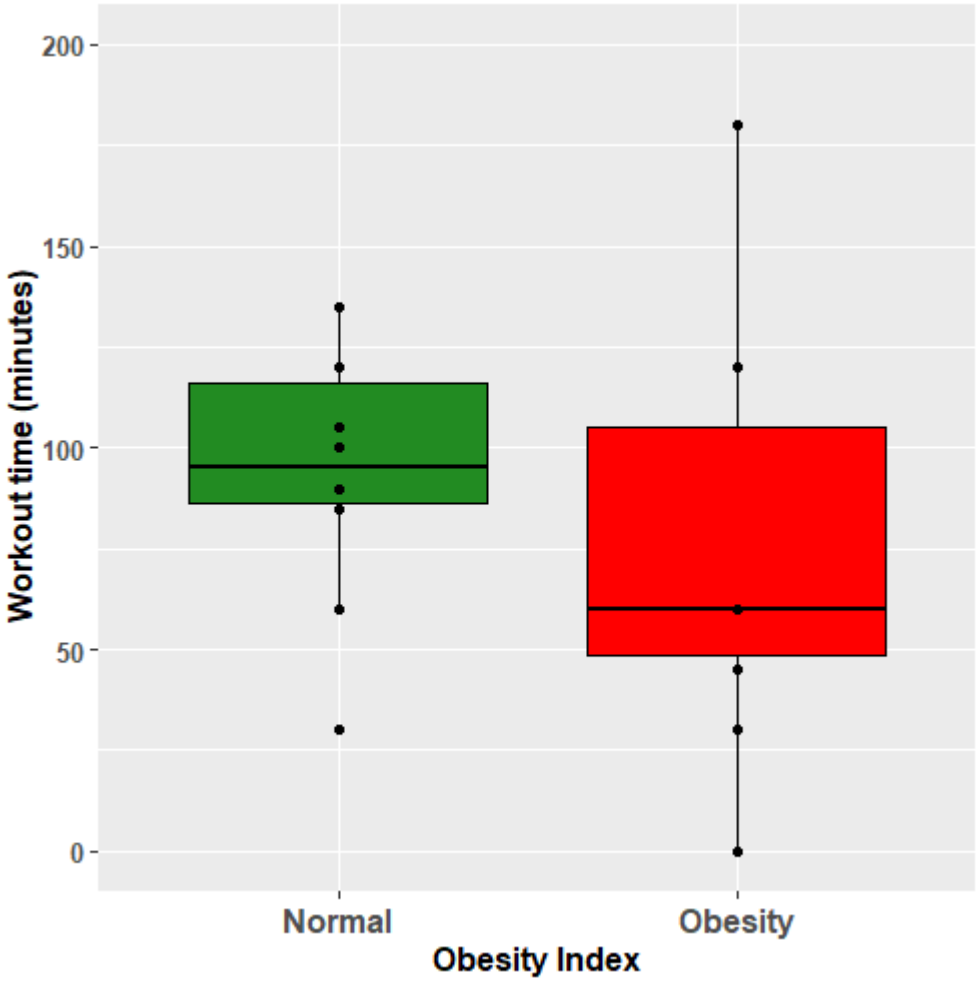


グラフセッティング

⑤点プロットの形式でデータを表記する

```
ggplot(obesity_data_renamed, aes(x = Obesity, y = Workout, fill = Obesity)) +
  geom_boxplot(color = 'black', outlier.shape = NA) +
  scale_fill_manual(values = c('forestgreen', 'red')) +
  xlab('Obesity Index') +
  ylab('Workout time (minutes)') +
  theme(axis.title.x = element_text(size = 12, face = 'bold'),
        axis.title.y = element_text(size = 12, face = 'bold'),
        axis.text.x = element_text(size = 12, face = 'bold'),
        axis.text.y = element_text(size = 10, face = 'bold')) +
  scale_y_continuous(limits = c(0,200)) +
  theme(legend.position = "none") +
  geom_point(color = "black")
```

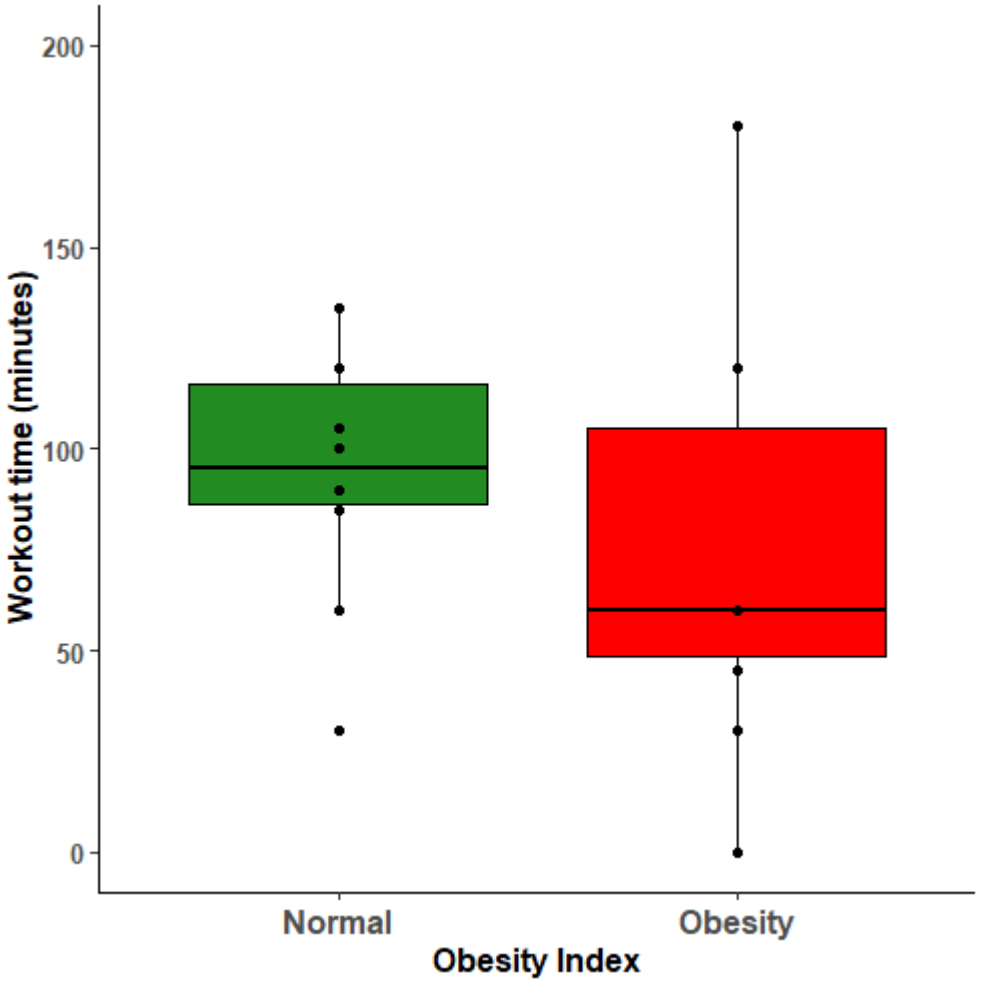
黒い点の形式ですべてのデータを表記



グラフセッティング

⑥グラフの背景を削除する

```
ggplot(obesity_data_renamed, aes(x = Obesity, y = Workout, fill = Obesity)) +
  geom_boxplot(color = 'black', outlier.shape = NA) +
  scale_fill_manual(values = c('forestgreen', 'red')) +
  xlab('Obesity Index') +
  ylab('Workout time (minutes)') +
  # Modified setting #
  theme(axis.title.x = element_text(size = 12, face = 'bold'),
        axis.title.y = element_text(size = 12, face = 'bold'),
        axis.text.x = element_text(size = 12, face = 'bold'),
        axis.text.y = element_text(size = 10, face = 'bold'),
        legend.position = "none",
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black")) +
  ###################
  scale_y_continuous(limits = c(0,200)) +
  geom_point(color = "black")
```

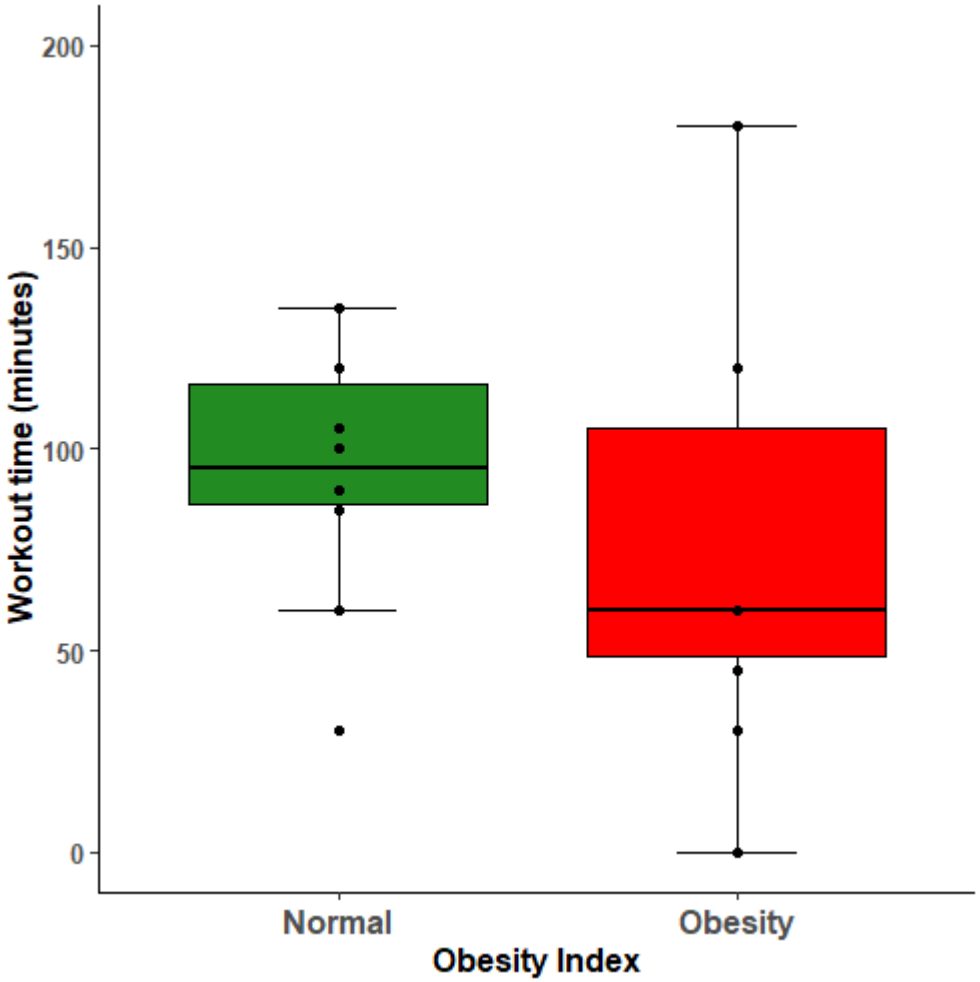


グラフセッティング

⑦エラーバーにひげを追加する

```
ggplot(obesity_data_renamed, aes(x = Obesity, y = Workout, fill = Obesity)) +
  # Modified setting #
  stat_boxplot(geom= 'errorbar' , width = 0.3, position = position_dodge(width = 0.75)) +
  ###################
  geom_boxplot(color = 'black', outlier.shape = NA) +
  scale_fill_manual(values = c('forestgreen', 'red')) +
  xlab('Obesity Index') +
  ylab('Workout time (minutes)') +

  theme(axis.title.x = element_text(size = 12, face = 'bold'),
        axis.title.y = element_text(size = 12, face = 'bold'),
        axis.text.x = element_text(size = 12, face = 'bold'),
        axis.text.y = element_text(size = 10, face = 'bold'),
        legend.position = "none",
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black")) +

scale_y_continuous(limits = c(0,200)) +
  geom_point(color = "black")
```

箱ひげ図を作る前に
エラーバーを作る

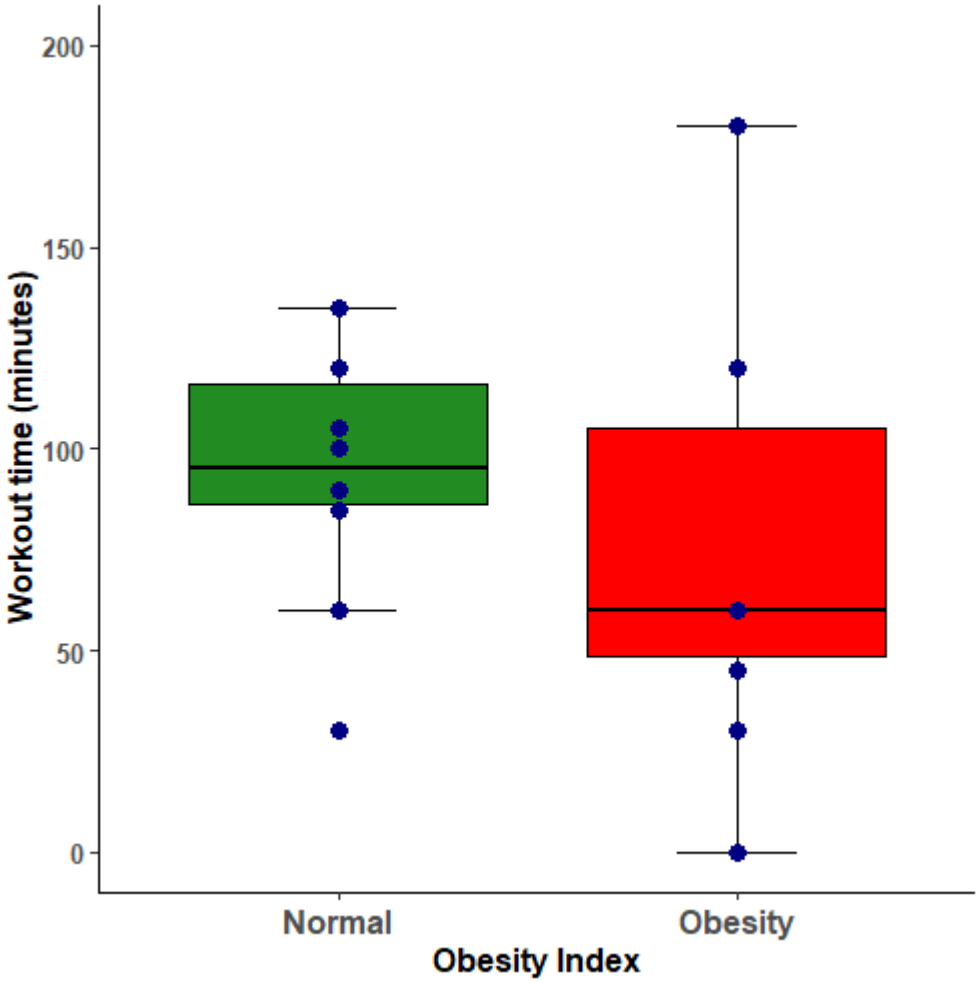


グラフセッティング

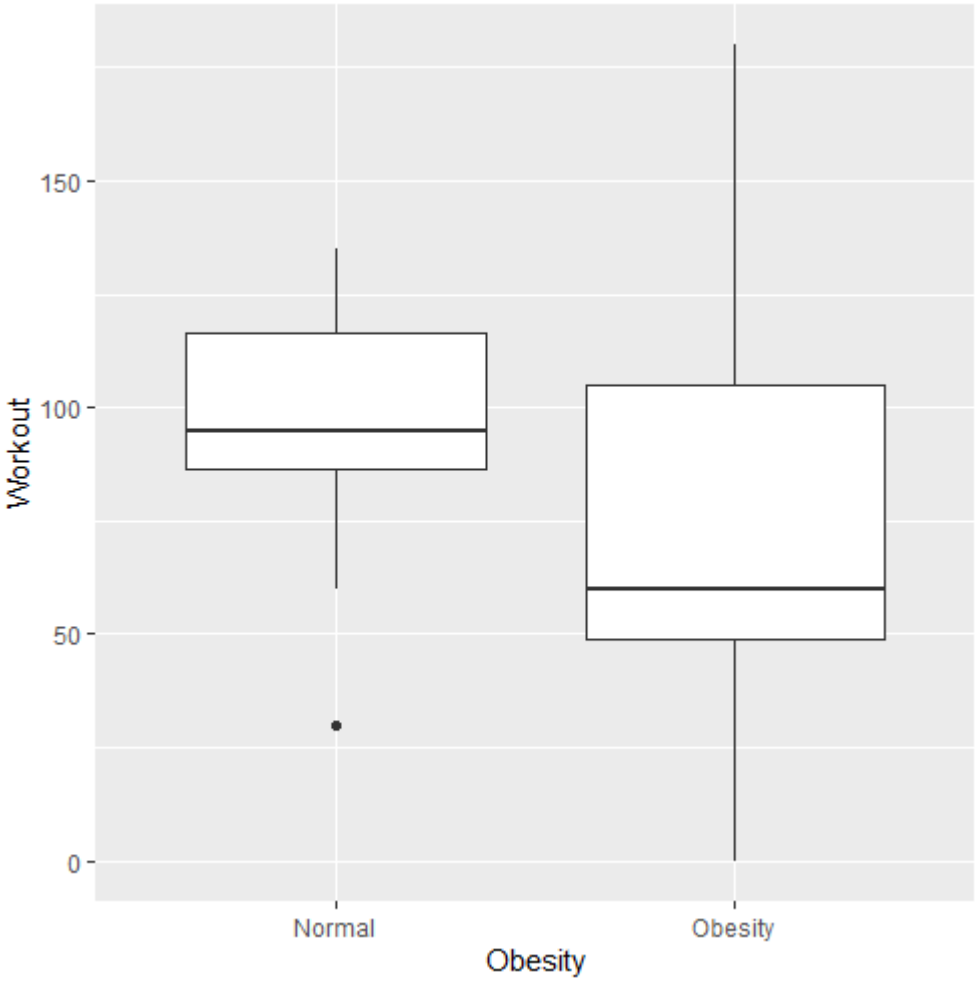
⑧点プロットを見やすくする

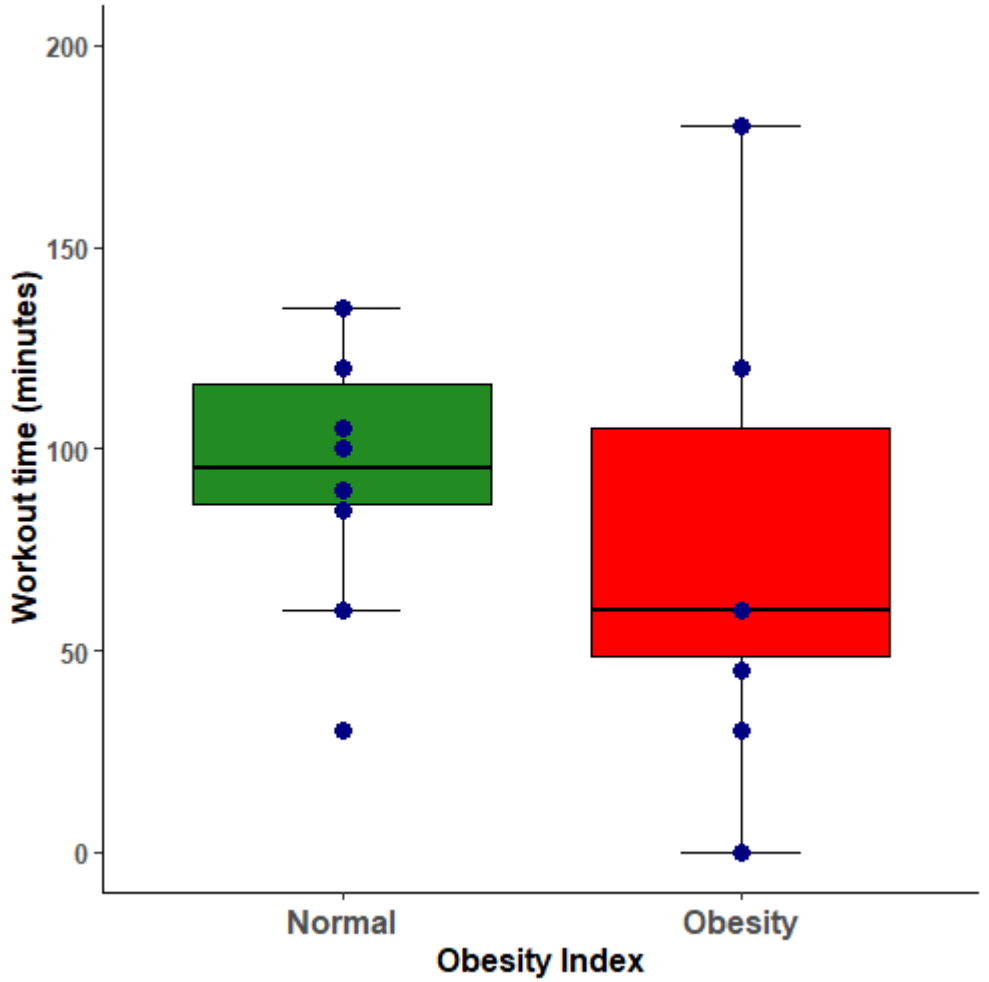
```
ggplot(obesity_data_renamed, aes(x = Obesity, y = Workout, fill = Obesity)) +
  stat_boxplot(geom= 'errorbar' , width = 0.3, position = position_dodge(width = 0.75)) +
geom_boxplot(color = 'black', outlier.shape = NA) +
  scale_fill_manual(values = c('forestgreen', 'red')) +
  xlab('Obesity Index') +
  ylab('Workout time (minutes)') +

  theme(axis.title.x = element_text(size = 12, face = 'bold'),
        axis.title.y = element_text(size = 12, face = 'bold'),
        axis.text.x = element_text(size = 12, face = 'bold'),
        axis.text.y = element_text(size = 10, face = 'bold'),
        legend.position = "none",
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        panel.background = element_blank(),
        axis.line = element_line(colour = "black")) +

  scale_y_continuous(limits = c(0,200)) +
  # Modified setting #
  geom_point(color = "navy", size = 3)
  ###################
```

色をネイビー、サイズを3に変更



結果


150
100
50
0
Workout
Normal
Obesity
Obesity


200
150
100
50
0
Workout time (minutes)
Normal
Obesity
Obesity Index