

# Quiz 5

Name:

## titanic data

For the questions today, you will be using data for passengers from the Titanic. Here is the start of this dataset, called `titanic`:

```
titanic <- read.csv("../data/titanic.csv", as.is = TRUE) %>%  
  select(Survived, Pclass, Age, Name, Sex, Fare)  
head(titanic)
```

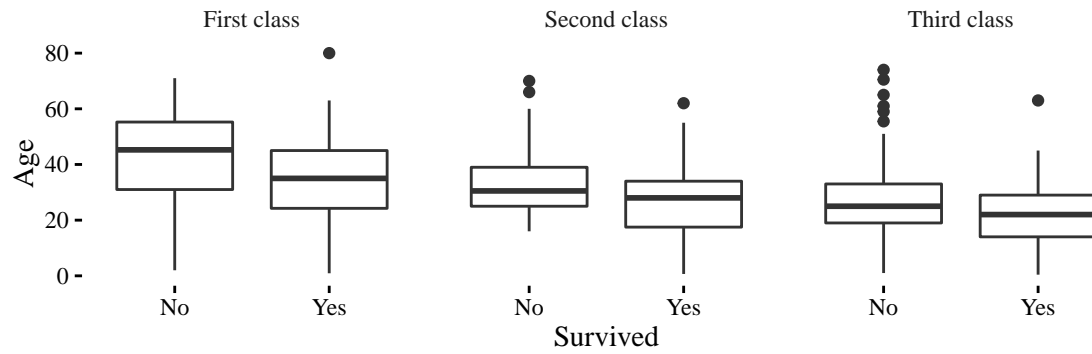
```
##   Survived Pclass Age                               Name  
## 1         0      3  22                               Braund, Mr. Owen Harris  
## 2         1      1  38 Cumings, Mrs. John Bradley (Florence Briggs Thayer)  
## 3         1      3  26                               Heikkinen, Miss. Laina  
## 4         1      1  35 Futrelle, Mrs. Jacques Heath (Lily May Peel)  
## 5         0      3  35                               Allen, Mr. William Henry  
## 6         0      3  NA                               Moran, Mr. James  
##      Sex   Fare  
## 1   male 7.2500  
## 2 female 71.2833  
## 3 female  7.9250  
## 4 female 53.1000  
## 5   male  8.0500  
## 6   male  8.4583
```

1.–6.

Match the output to the numbers written in comments at the end of code lines. Each correct answer will show the output if you printed it out as of that line of code (i.e., you've run that line of the code with the exception of any `%>%` or `+` at the line end). Each number will have one and only one answer. Write your answer beside the question number in the code. For any answers that are dataframes, I'm only showing the head of the dataframe.

```
toplot_1 <- filter(titanic, !is.na(Age)) %>%                                # 1.  
  select(-Name, -Sex, -Fare) %>%                                           # 2.  
  mutate(Survived = factor(Survived,  
                            levels = c(0, 1),  
                            labels = c("No", "Yes")),  
         Pclass = factor(Pclass, levels = c(1, 2, 3),  
                           labels = c("First class", "Second class",  
                                       "Third class"))))                  # 3.  
my_plot <- ggplot(toplot_1, aes(x = Survived, y = Age)) +  
  geom_boxplot() +                                                         # 4.  
  theme_tufte() +                                                         # 5.  
  facet_grid(. ~ Pclass)                                                  # 6.
```

a.



b.

##	Survived	Pclass	Age	Name
## 1	0	3	22	Braund, Mr. Owen Harris
## 2	1	1	38	Cummings, Mrs. John Bradley (Florence Briggs Thayer)
## 3	1	3	26	Heikkinen, Miss. Laina
## 4	1	1	35	Futrelle, Mrs. Jacques Heath (Lily May Peel)
## 5	0	3	35	Allen, Mr. William Henry
## 6	0	3	NA	Moran, Mr. James

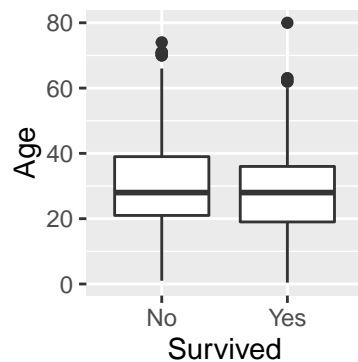
  

##	Sex	Fare
## 1	male	7.2500
## 2	female	71.2833
## 3	female	7.9250
## 4	female	53.1000
## 5	male	8.0500
## 6	male	8.4583

c.

##	Survived	Pclass	Age
## 1	No	Third class	22
## 2	Yes	First class	38
## 3	Yes	Third class	26
## 4	Yes	First class	35
## 5	No	Third class	35
## 6	No	First class	54

d.



e.

##	Survived	Pclass	Age
## 1	0	3	22
## 2	1	1	38
## 3	1	3	26
## 4	1	1	35
## 5	0	3	35
## 6	0	1	54

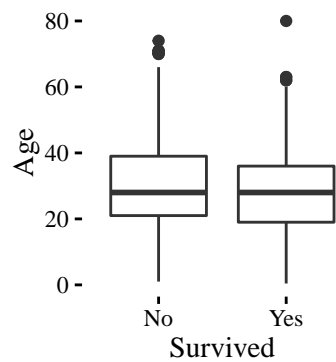
f.

##	Survived	Pclass	Age	Name
## 1	0	3	22	Braund, Mr. Owen Harris
## 2	1	1	38	Cumings, Mrs. John Bradley (Florence Briggs Thayer)
## 3	1	3	26	Heikkinen, Miss. Laina
## 4	1	1	35	Futrelle, Mrs. Jacques Heath (Lily May Peel)
## 5	0	3	35	Allen, Mr. William Henry
## 6	0	1	54	McCarthy, Mr. Timothy J

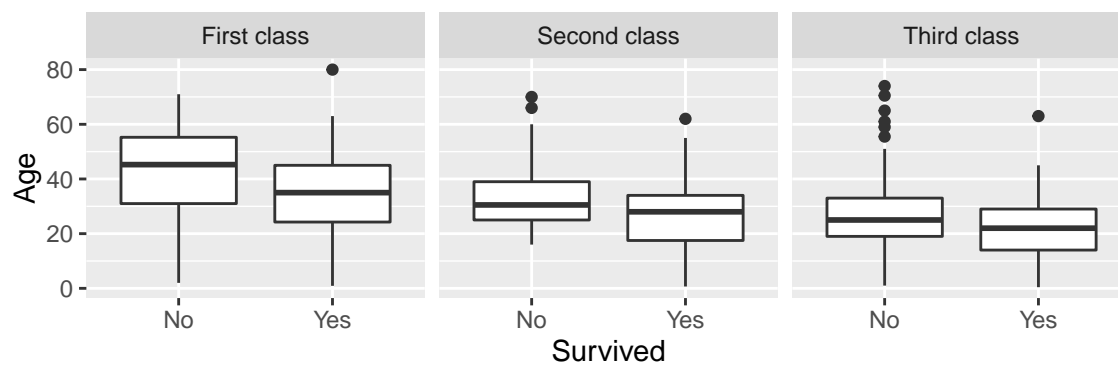
  

##	Sex	Fare
## 1	male	7.2500
## 2	female	71.2833
## 3	female	7.9250
## 4	female	53.1000
## 5	male	8.0500
## 6	male	51.8625

g.



h.



Answers: 1. f., 2. e., 3. c., 4. d., 5. g., 6. a.

7.–8.

You create the following function:

```
clean_data <- function(df, age){  
  df_subset <- subset(df, Age == age) %>%  
    filter(!is.na(Age)) %>%  
    select(Survived, Sex, Age, Name)  
  return(df_subset)  
}
```

What would you get if you ran the following code? (Only one answer each; please write the letter by the number in the code.)

```
df_35 <- clean_data(df = titanic, age = 35)  
head(df_35, 3) # 7.  
with(df_35, table(Sex, Survived)) # 8.
```

a.

```
##   Survived   Sex Age                                     Name  
## 1         0  male 22                               Braund, Mr. Owen Harris  
## 2         1 female 38 Cumings, Mrs. John Bradley (Florence Briggs Thayer)  
## 3         1 female 26                               Heikkinen, Miss. Laina
```

b.

```
##   Survived   Sex Age                                     Name  
## 1         1 female 35 Futrelle, Mrs. Jacques Heath (Lily May Peel)  
## 2         0  male 35                               Allen, Mr. William Henry  
## 3         0  male 35                               Fynney, Mr. Joseph J
```

c.

```
##           Survived  
## Sex           0 1  
##  female 0 8  
##  male   7 3
```

d.

```
##   Survived Pclass Age                                     Name   Sex  
## 1         1     1  35 Futrelle, Mrs. Jacques Heath (Lily May Peel) female  
## 2         0     3  35           Allen, Mr. William Henry   male  
## 3         0     2  35           Fynney, Mr. Joseph J     male  
##   Fare  
## 1 53.10  
## 2  8.05  
## 3 26.00
```

e.

```
## Sex
## female  male
##      8    10
```

```
## Survived
##  0  1
##  7 11
```

**Answers:** 7. b., 8. c.