S2S Lab 4 Task Solutions

1 Welcome!

2 Qualitative Data

2.1 Tables

The levels of `ease` are automatically presented in alphabetical order. Can you reorder the labels so that they go in the order `"Easy"`, `"Difficult"` and `"Impossible"`?

```
EPIDURAL$ease <- factor(x = EPIDURAL$ease,
                        levels = c("Easy", "Difficult", "Impossible"))
levels(EPIDURAL$ease)
```

```
## [1] "Easy"       "Difficult"  "Impossible"
```

The function `levels()` can be used to see the levels of a factor variable, as well as the order they are stored in.

We can update the frequency table `epi_freq` so that the different levels of `ease` now appear in a sensible order.

```
epi_freq <- table(EPIDURAL$ease)
epi_freq
```

```
##
##       Easy  Difficult Impossible
##         57         20          8
```

The `TITANIC3` data set from the `PASWR2` package contains information on the survival status of passengers on the Titanic. Load this data set into your Environment tab and create a frequency table showing the number of passengers who survived and the number who did not.

Remember to explore the data set using the `help()` and `str()` functions. It might help to rename the levels of the `survived` variable so they are clear.

```
data("TITANIC3")

TITANIC3$survived <- factor(x = TITANIC3$survived, labels = c("No", "Yes"))

table(TITANIC3$survived)
```

```
##
##  No Yes
## 809 500
```

What proportion of passengers survived the Titanic sinking?

We can find the proportion of survivors by using the `prop.table()` function.

```
prop.table(table(TITANIC3$survived))
```

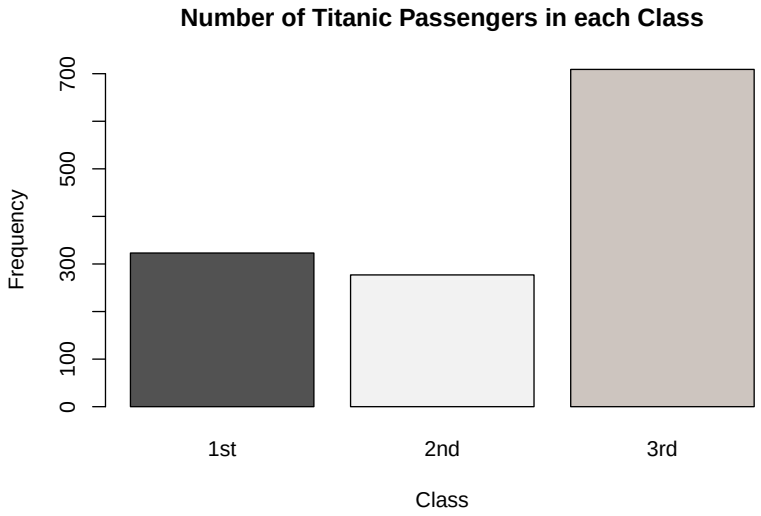
```
##
```

```
##       No      Yes
## 0.618029 0.381971
```

2.2 Barplots

Complete the code below to create a barplot which shows the number of passengers on the Titanic belonging to 1^{st}, 2^{nd} and 3^{rd} class respectively.

```
barplot(height = table(TITANIC3$pclass),
        col = c("gray32", "gray95", "seashell3"),
        main = "Number of Titanic Passengers in each Class",
        xlab = "Class",
        ylab = "Frequency")
```

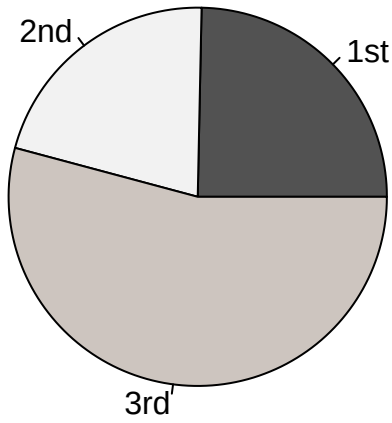


2.3 Dot Charts

2.4 Pie Charts

Create a pie chart showing the proportion of passengers on the Titanic in each class.

```
pie(x = table(TITANIC3$pclass),
    col = c("gray32", "gray95", "seashell3"),
    main = "Class of Titanic Passengers")
```



2.5 Bivariate Data

Create a two-way contingency table showing the number of passengers who survived and did not survive in each of the three classes (1^{st}, 2^{nd} and 3^{rd}).

```
table(TITANIC3$pclass, TITANIC3$survived)
```

```
##
##         No Yes
##   1st  123 200
##   2nd  158 119
##   3rd  528 181
```

How many 3^{rd} class passengers survived the sinking of the Titanic?

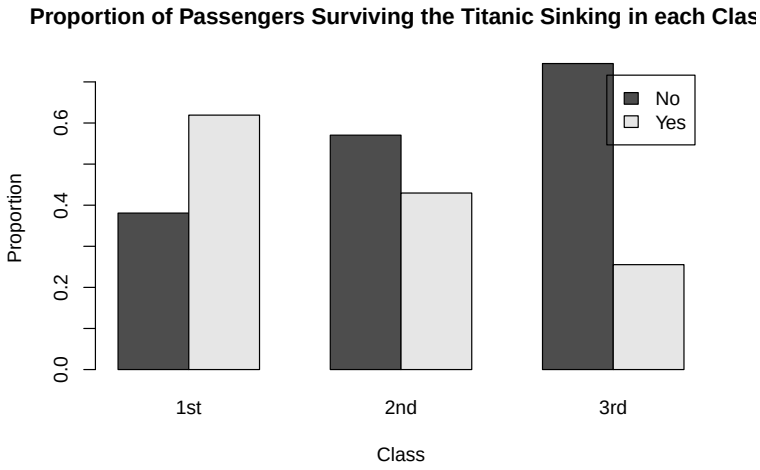
181

Create a barplot showing the proportion of passengers in each class who survived and the proportion who did not.

Make sure the bars are side-by-side, and that the "class" variable is shown along the x-axis.

```
surv_prop <- prop.table(table(TITANIC3$pclass, TITANIC3$survived), margin = 1)

barplot(height = t(surv_prop),
        legend.text = TRUE,
        beside = TRUE,
        main = "Proportion of Passengers Surviving the Titanic Sinking in each Class",
        xlab = "Class",
        ylab = "Proportion")
```



3 Quantitative Data

3.1 Stem and Leaf Plots

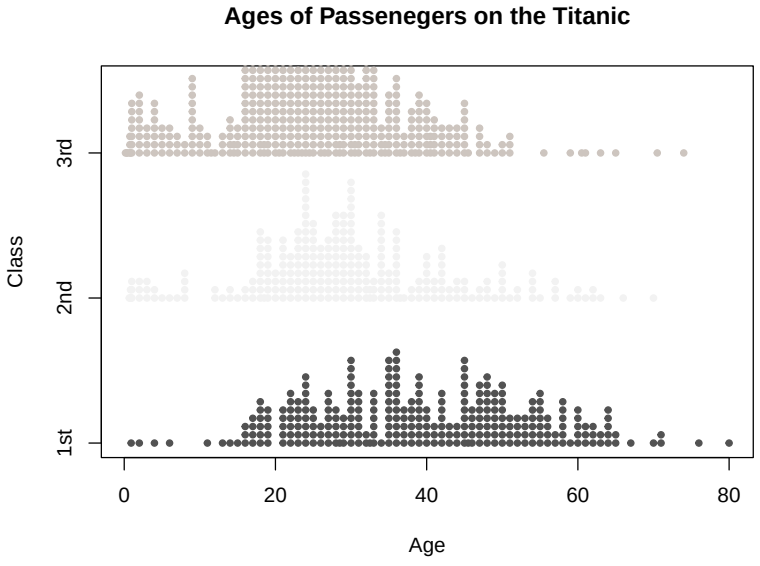
How would you describe the shape of the distribution of patients' weights?

Right-skewed

3.2 Strip Charts

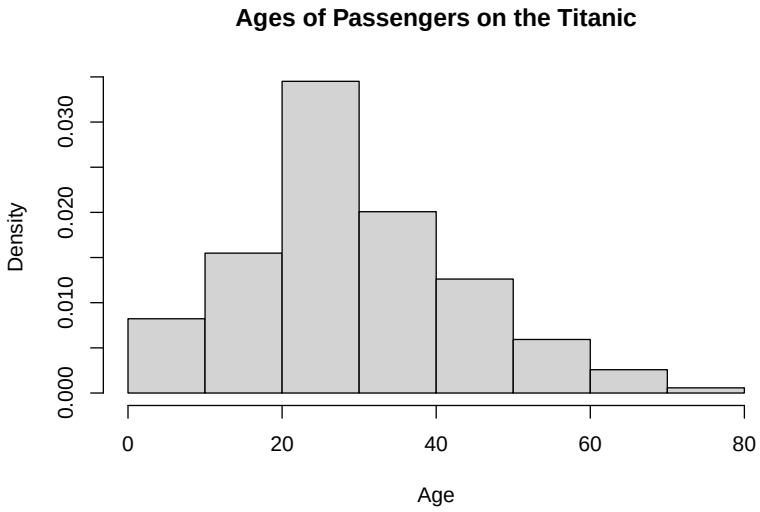
Complete the code below to create a stripchart showing the distribution of passenger ages in each class on board the Titanic.

```
stripchart(x = age ~ pclass,
           data = TITANIC3,
           method = "stack",
           pch = 20,
           col = c("gray32", "gray95", "seashell3"),
           main = "Ages of Passenegers on the Titanic",
           xlab = "Age",
           ylab = "Class")
```



3.3 Histograms

What single value would you provide to the `breaks =` argument to create the density histogram of passenger ages on the Titanic shown below?



If you instead wanted to provide the argument `breaks =` with a sequence generated using the `seq()` function, how would you do this?

Both chunks of code below will produce the same histogram shown above.

```
hist(x = TITANIC3$age,
     breaks = 8,
     freq = FALSE,
     main = "Ages of Passengers on the Titanic",
     xlab = "Age")
```

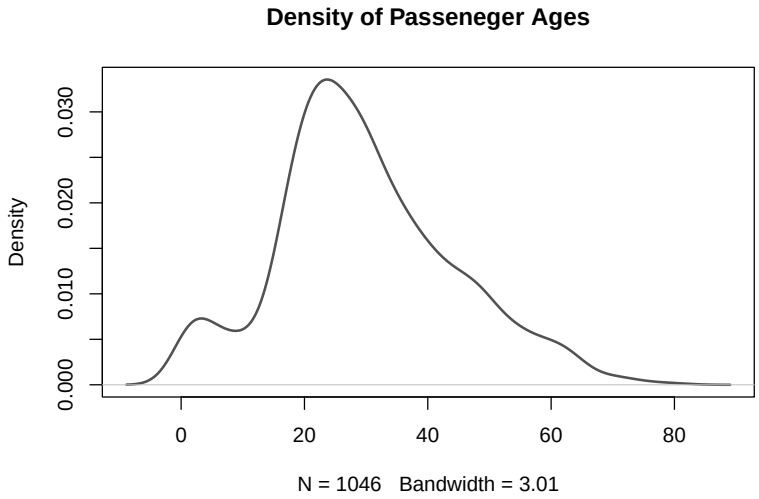
```
hist(x = TITANIC3$age,
     breaks = seq(from = 0, to = 80, by = 10),
     freq = FALSE,
     main = "Ages of Passengers on the Titanic",
     xlab = "Age")
```

3.4 Kernel Density Estimators

Create a plot only showing the kernel density estimate for distribution of passenger ages on the Titanic.

```
age_dens <- density(x = TITANIC3$age, na.rm = TRUE)

plot(age_dens, col = "gray32", lwd = 2,
     main = "Density of Passeneger Ages")
```



4 Comparing Samples

4.1 Comparing Histograms

How would you compare the two distributions shown in the histograms above?

- "Both distributions are skewed to the right."
- "The distribution for the hamstring stretch group is more extremely skewed than the distribution for the traditional sitting group."

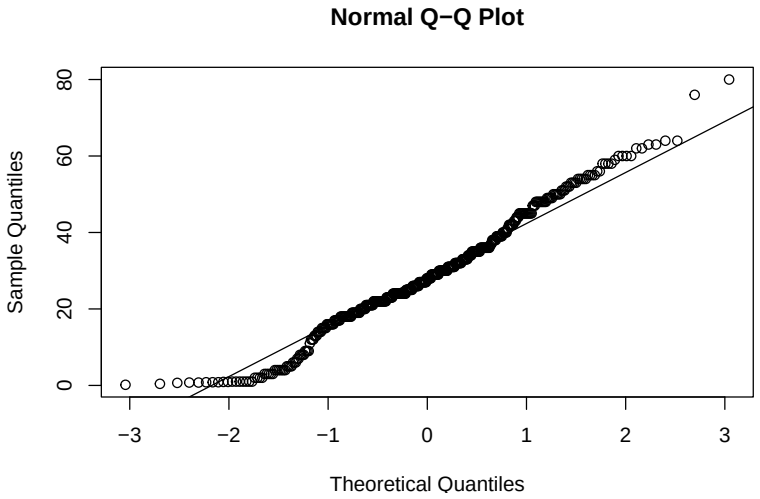
4.2 Comparing Density Estimates

4.3 QQ Plots

Create a new subset of the `TITANIC3` data frame that contains the ages of all passengers who survived the sinking. Plot the quantiles of the distribution of age against the quantiles of a standard normal distribution in a QQ plot.

```
age_survived <- subset(x = TITANIC3,
                       subset = (survived == "Yes"),
                       select = age)

qqnorm(y = age_survived$age)
qqline(y = age_survived$age)
```



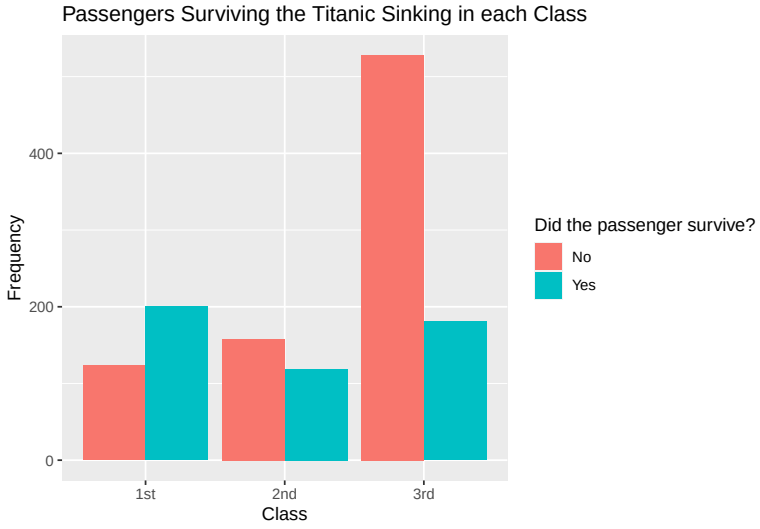
Do you think that the ages of passengers who survived the sinking of the Titanic follow a normal distribution?

No

5 ggplot2

5.1 Barplots

Complete the code below to create the barplot showing the number of passengers who survived and did not survive in each class from the `TITANIC3` data set beneath.



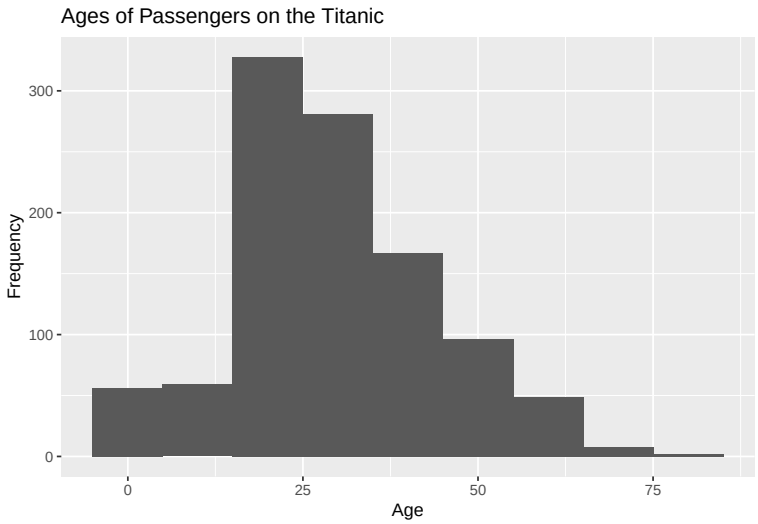

```
ggplot(data = TITANIC3) +
  geom_bar(aes(x = pclass, fill = survived), position = "dodge") +
  labs(title = "Passengers Surviving the Titanic Sinking in each Class",
       x = "Class", y = "Frequency") +
  guides(fill = guide_legend("Did the passenger survive?"))
```

5.2 Histograms

Create a histogram, using ggplot2, showing the ages of all passengers on the Titanic. Label your plot appropriately and consider using different binwidths to explore the shape of the

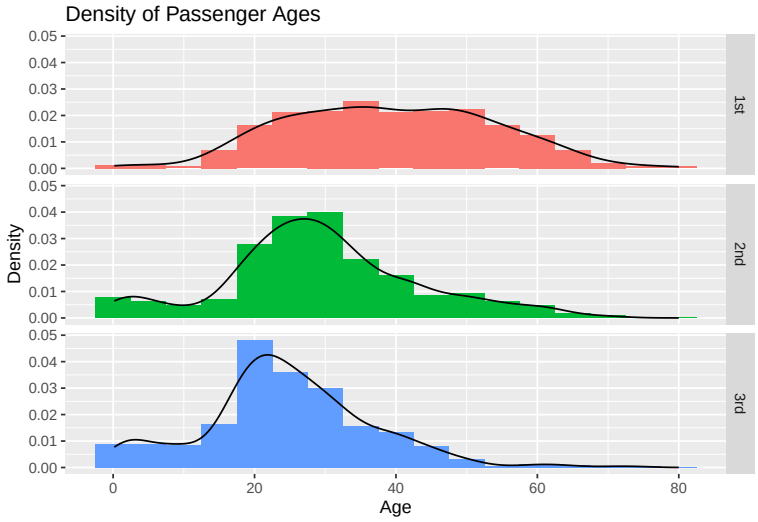
distribution.

```
ggplot(data = TITANIC3) +
  geom_histogram(aes(x = age), binwidth = 10) +
  labs(title = "Ages of Passengers on the Titanic",
       x = "Age", y = "Frequency")
```



5.3 Kernel Density Estimators

Complete the code below to create the following histogram and density plots which show the distribution of passenger age for each class of passengers on board the Titanic.




```
ggplot(data = TITANIC3) +
  geom_histogram(aes(x = age, y = after_stat(density), fill = pclass),
                 binwidth = 5) +
  geom_density(aes(x = age)) +
  facet_grid(pclass ~ .) +
  labs(title = "Density of Passenger Ages",
       x = "Age", y = "Density") +
  guides(fill = "none")
```