**Lab Assignment #4 (Due: 11:00am, Thu, Sep 6)**

**Write Your Name Here**

**1. Old Faithful Geyser Eruptions** (**Data: OldFaithful**)

Millions of people from around the world flock to Yellowstone Park in order to watch eruptions of the Old Faithful geyser. (<https://www.nps.gov/yell/index.htm>)

How long does a person usually have to wait between eruptions, and has the timing changed over the years? In particular, scientists have investigated whether a 1998 earthquake (see below) lengthened the time between eruptions at Old Faithful.

**The March 5, 1998 Coso Earthquake** <http://service.scedc.caltech.edu/ftp/ca.earthquakes/specialrpt/1998/980305.html>

March 6, 1998: The largest earthquake in southern California since March 1997 occurred on Thursday night, March 5, at 9:47 p.m. PST. It was a M5.2 earthquake located in the Owens Valley, 17 east-northeast of the small town of Little Lake. As of 2:30 p.m. on Friday March 6, SCSN/TriNet (the Southern California Seismic Network component of TriNet) had recorded 164 aftershocks to this earthquake. The largest 3 are M4.8 at 9:49, M4.6 at 9:54 p.m., and M4.5 at 11:36 p.m.

The Owens Valley is an extensional regime, but the predominant focal mechanisms are strike-slip. This earthquake is also strike slip with planes striking northwest and northeast. The aftershocks strike northeast, suggesting that the causative fault is left-lateral strike-slip on the northeast-striking plane

The data set contains the inter-eruption times (i.e. number of minutes between two consecutive eruptions) for all eruptions occurred between 6:00 a.m. and midnight on Aug 1 – 8 in 1978 and also in 2003.

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Minutes | waiting time (in minutes) from one eruption to the next eruption |
| Year | year of eruption (Year1978 or Year2003) |

* First, import the data file “**OldFaithful.csv**” into RStudio.
* Then, attach the data file

> attach(OldFaithful)

**(a)** Let’s compare the waiting times (i.e. inter-eruption times) between year 1978 (before the Coso Earthquake) and year 2003 (after the Coso Earthquake).

* Draw a box plot (not bar plot) of waiting times in each year.
* Add a title (your choice) in the plot and a label for the variable

## The box will appear vertically

> boxplot(Minutes~Year, horizontal=T, main="your title here!!)", xlab="Waiting Time in minutes" )

OR

## The box will appear horizontally

> boxplot(Minutes~Year main="your title here!!)", ylab="Waiting Time in minutes")

* Paste the R commands inside a text box.
* Paste the plot here. Resize the plot if it is too large or too small.

**(b)** Calculate the mean waiting time, the five-number summary of waiting time times for each year so that we can compare them.

* Paste the R command and the results here inside a text box.

**(c)** Scientists have investigated whether a 1998 earthquake lengthened the time between eruptions at Old Faithful. Do you think that waiting times (inter-eruption times) tend to be longer or shorter in one year than the other year? Explain why or why not using the results in previous parts.

**(d)** Calculate the standard deviation and the IQR of waiting times for each year so that we can compare them**.**

* Paste the R command and the result here inside a text box.

**(e)** Park rangers and visitors would also appreciate if Old Faithful was as reliable as its name implies. Do you think waiting times tend to be more consistent in one year than the other year? Explain why or why not using the results in previous parts.

**Note: Detach the file when you are done**

> detach(OldFaithful)

**2. The Titanic (Data:** **Titanic)**

On April 15, 1912, on her maiden voyage, the Titanic collides with an iceberg and sank. The ship was luxurious but did not have enough lifeboats for the 2,224 passengers and crew. As a result of the collision, 1,502 people died. The ship had three classes of passengers. The level of luxury and the price of the ticket varied with the class, with first class being the most luxurious.

* First, import the data file “**Titanic.csv**” into RStudio.
* Then, attach the data file

**(a)** We compared the agespassengers among the First, Second, and Third classes. Draw a (side-by-side) boxplot so that we can compare the ages of three groups.

* Paste the R command inside a text box.
* Paste the plot here.

**(b)** We can also calculate the median ages (or any statistics) of passengers for all combinations of ticket class and gender.

* Calculate the median age for each group. Paste the R command inside a text box.

> tapply(Age, Class:Gender, median)

* Report the median ages in the table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | **Ticket Class** | | |
| **First Class** | **Second Class** | **Third Class** |
| **Gender** | **Female** |  |  |  |
| **Male** |  |  |  |

**(c)** Similarly, draw (side-by-side) boxplots for ages for all combinations of ticket class and gender (six groups).

> boxplot(Age ~ Class:Gender)

* Paste the R command inside a text box.
* Paste the plot here.

**(d)** Tell me what you see in the plot in part (c). (There is no right or wrong answer,)