Homework 01 - Displaying Data

Problem 1

Part a)

```
# This command reads in a comma-separated-values file at the following url and stores it in

mosquitoes<-read.csv("http://math.arizona.edu/~jwatkins/mosquitoes.csv")
```

Part b)

Give the five number summary of the life span of both types of mosquitoes.
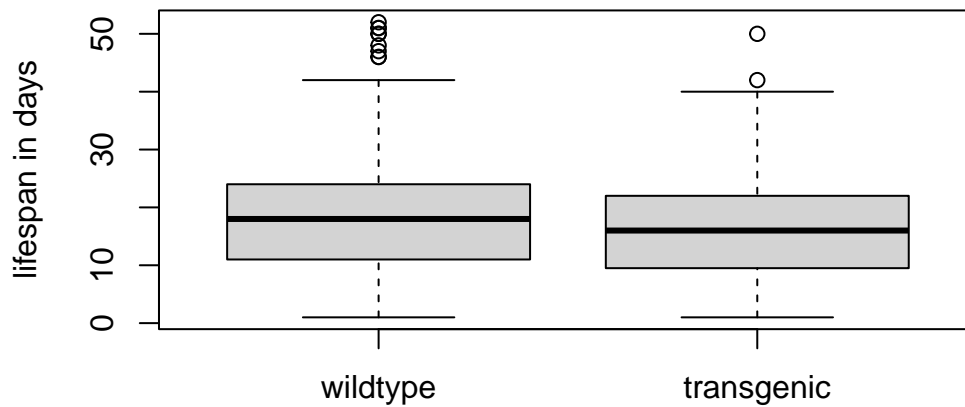
```
summary(mosquitoes)
```

```
    wildtype          transgenic
 Min.   : 1.00    Min.   : 1.00
 1st Qu.:11.00    1st Qu.: 9.50
 Median :18.00    Median :16.00
 Mean   :20.78    Mean   :16.55
 3rd Qu.:24.00    3rd Qu.:22.00
 Max.   :52.00    Max.   :50.00
 NA's   :11
```

Part c)

Give side by side box plots of the life span of both types of mosquitoes.

```
boxplot(mosquitoes, ylab="lifespan in days")
```

Part d)

Graph the empirical survival functions for both types of mosquitoes in one graph. In the code block below make sure to complete the following tasks:

- Annotate the survival function where indicated
- Fill in the appropriate data names (notice the formatting of calling a column from a dataset)

We start by creating two new vectors to do this: `wildtype` and `transgenic` . To overlay the empirical survival functions use the command `par(new=TRUE)`. Be sure to give the same limits for the values on each of the axes and use different colors for each mosquito type.

```r
library(stats)

survival_function <- function(data) {
  ecdf_func <- ecdf(data)

  # EXPLAIN why 1-ecdf_func(t) is appropriate here. Hint: What is true about the survival fu
  survival_func <- function(t) {
    1 - ecdf_func(t)
  }
}
```

```
  return(survival_func)
}


wildtype <- survival_function(mosquitoes$wildtype)

transgenic <- survival_function(mosquitoes$transgenic)

plot(wildtype, xlim=c(0,55),ylim=c(0,1),
    main="Mosquito Lifespan",xlab="Days Survived",ylab="Surviving Fraction",col="blue")

par(new=TRUE)

plot(transgenic, xlim=c(0,55),ylim=c(0,1),
    main="Mosquito Lifespan",xlab="Days Survived",ylab="Surviving Fraction",col="red")

legend("topright",legend=c("wildtype","transgenic"),fill=c("blue","red"))
```
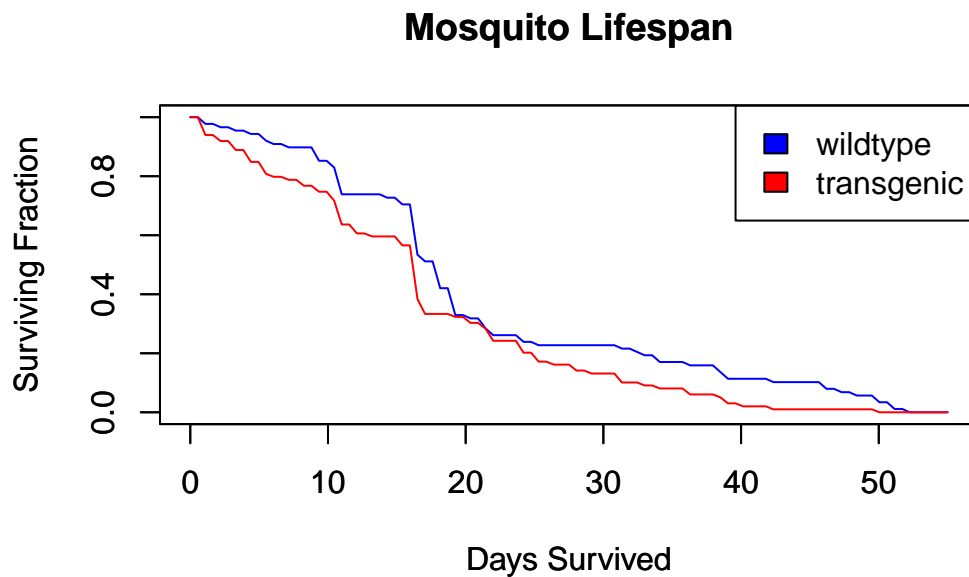

Part e)

Give the Q-Q plot of the two types of mosquitoes. Indicate the median and the first and third quartiles on the graph.
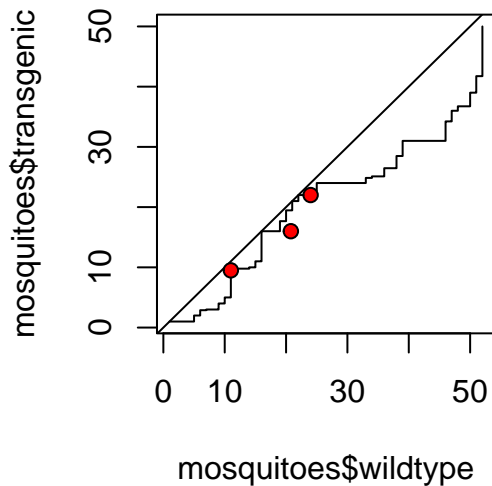
What is a qqplot? This type of plot matches up the various quantiles of the two data sets. For example, one of the points on the qqplot will be (median of wildtype data, median of transgenic type data).

```
qqplot(mosquitoes$wildtype, mosquitoes$transgenic ,type="s")
```

```
abline(a=0,b=1)
```

```
XXX = c( 11.00 , 20.78 , 24.00 )
YYY = c( 9.5 , 16.00 , 22.00 )
```

```
points(XXX,YYY,pch = 21,bg = "red")
```


Part f)

One genotype of mosquito lives longer, on average, than the other. Explain how this can be seen in the boxplots, in the survival function and on the Q-Q plot.

Boxplots: all quartiles are higher for wildtype mosquitoes, as is the max.

Survival function: the blue line for wildtype is consistently higher than the red line for transgenic

Q-Q plot: all the red dots are below the diagonal line, which means the wildtype mosquitoes consistently live longer.

Part g)

Before submitting, clean up this document. Anywhere in the code where there are NOTES and unnecessary information, remove those. You can leave in comments that you find necessary but think about making this document as nice as possible for a classmate to read. Once you have done so, render the document and upload to Gradescope.

Make sure to tag your pages in Gradescope! This means that each problem should have the appropriate pages associated with it.