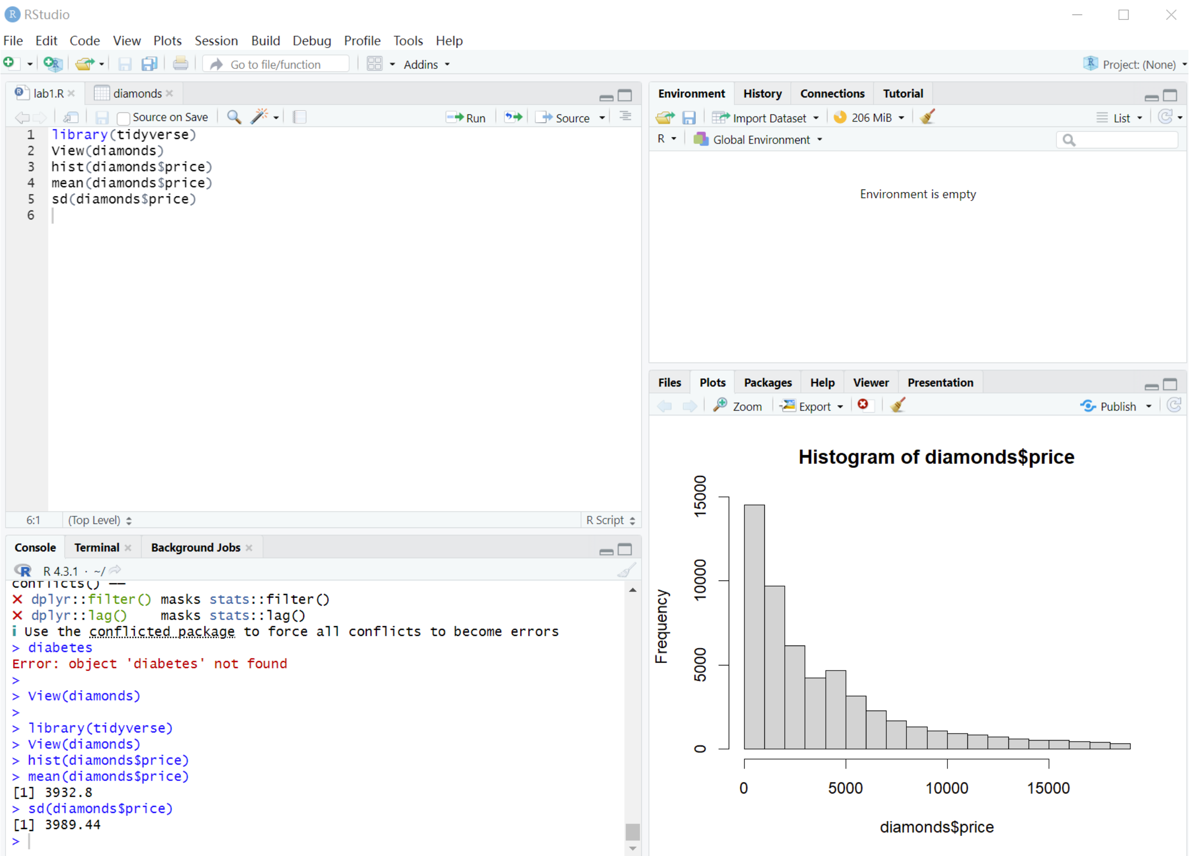
Stat212 Lab 1-Simulation Study with Diamonds

NAME: Yiran Hu NETID: yiranhu3

Q1.

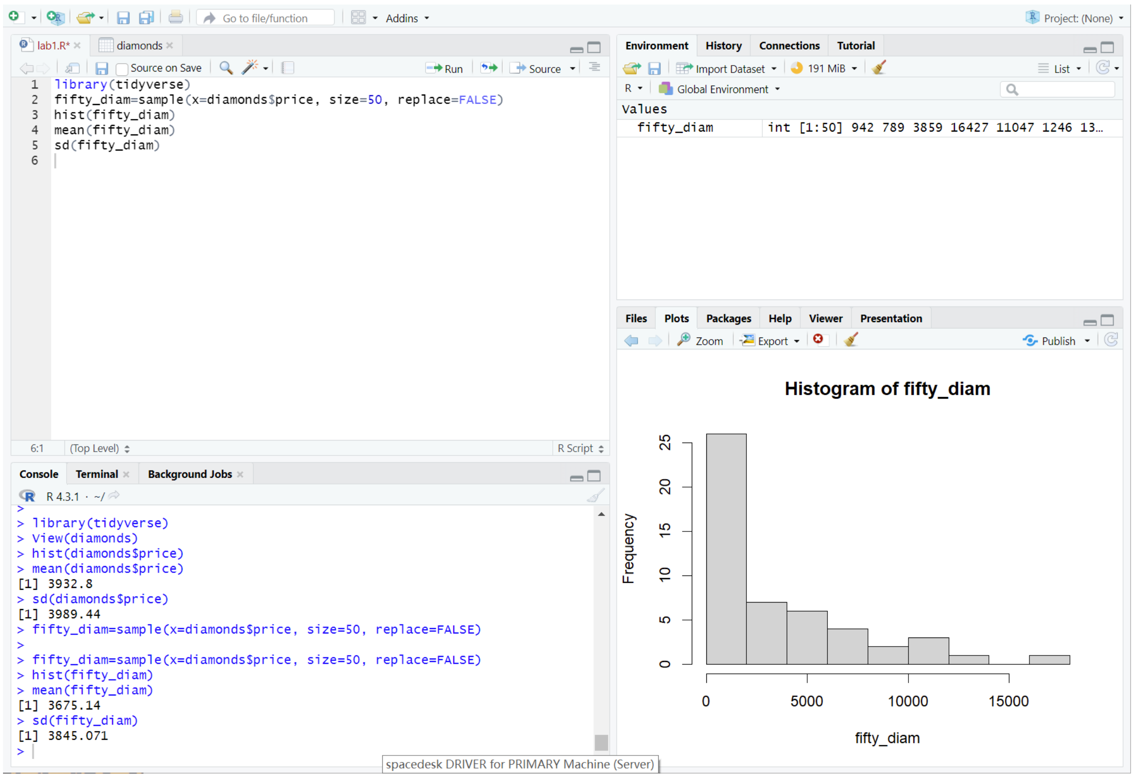
1. histogram



2. mean_1=3932.8, standard_deviation_1=3989.44
3. Right skewed

Q2.

1. histogram



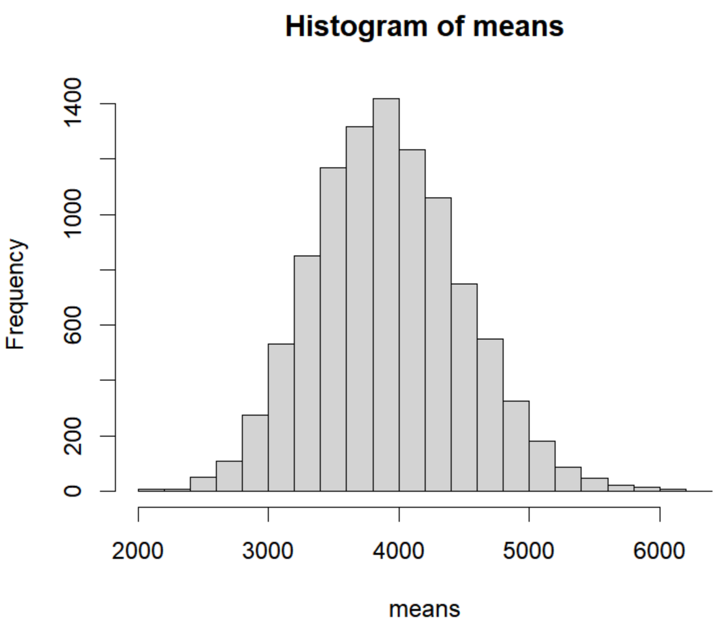
2. mean_2=3675.14, standard_deviation_2=3845.071

3. absolute error(mean)=|mean_1-mean_2|=257.66

4. absolute error(SD)=|standard_deviation_1-standard_deviation_2|=144.369

Q3.

1. histogram



2. code

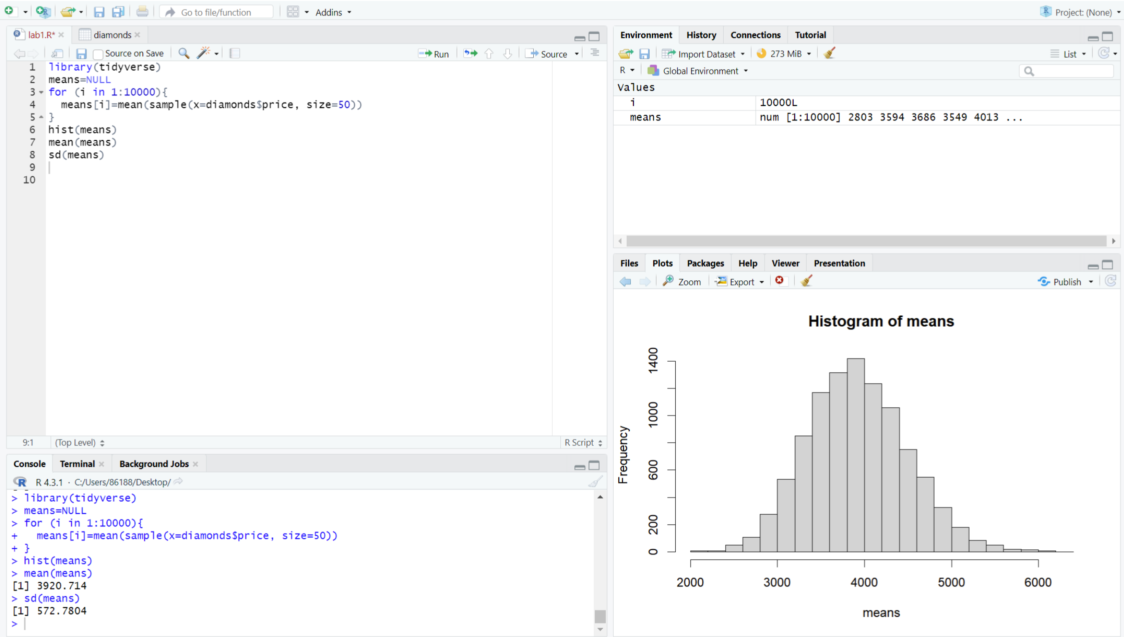
```
library(tidyverse)
means=NULL
for (i in 1:10000){
  means[i]=mean(sample(x=diamonds$price, size=50))
}
hist(means)
```

3. I would describe it as symmetric.

   By the Central Limit Theorem, even though the price is left skewed, the distribution of sample means will become normally distributed if the sample size generating those sample means is large enough. And the "Histogram of diamonds$price" seems to be highly right skewed and has a long tail, then the sample size need to be large for the "means" to be normally distributed. The sample size is 50, so the histogram is a little right skewed.

Q4.

1. standard_deviation_4=572.7804



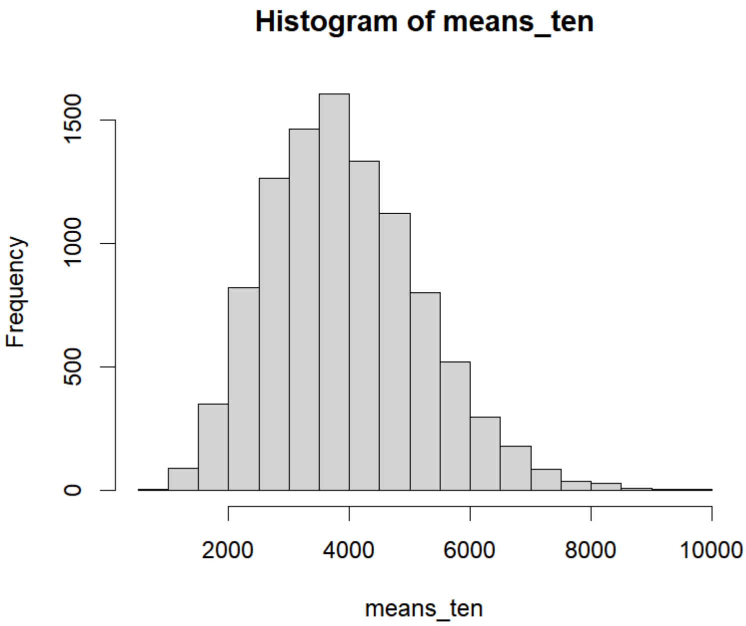
2. Standard Error,

   true value=population SD divided by the square root of the sample size

$$true\ value = \frac{standard\_deviation\_1}{\sqrt{size}} = \frac{3989.44}{\sqrt{50}} = 564.192$$

Q5.

1. histogram



2. code

```
library(tidyverse)
means_ten=NULL
for (i in 1:10000){
  means_ten[i]=mean(sample(x=diamonds$price, size=10))
}
hist(means_ten)
sd(means_ten)
```

3. shape

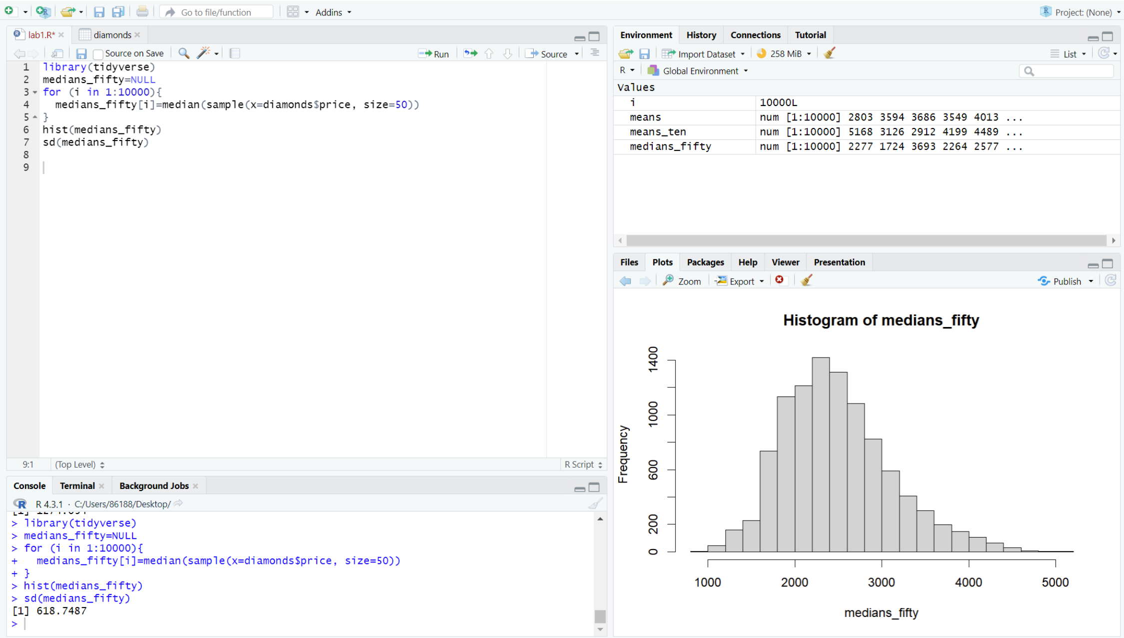
I would say it is right skewed. Because the "Histogram of diamonds$price" seems to be highly right skewed and has a long tail, so sample size need to be large enough (in lecture we say it need to be larger than 100) for "means" to be normally distributed. The sample size is 10, by Central Limit Theorem, which is not large enough to generate a normal distribution, so "Histogram of means_ten" looks right skewed.

4. Higher,

```
> library(tidyverse)
> means=NULL
> for (i in 1:10000){
+   means[i]=mean(sample(x=diamonds$price, size=50))
+ }
> hist(means)
> mean(means)
[1] 3920.714
> sd(means)
[1] 572.7804
> library(tidyverse)
> means_ten=NULL
> for (i in 1:10000){
+   means_ten[i]=mean(sample(x=diamonds$price, size=10))
+ }
> hist(means_ten)
> sd(means_ten)
[1] 1274.654
> 
```

Q6.

1. histogram & code



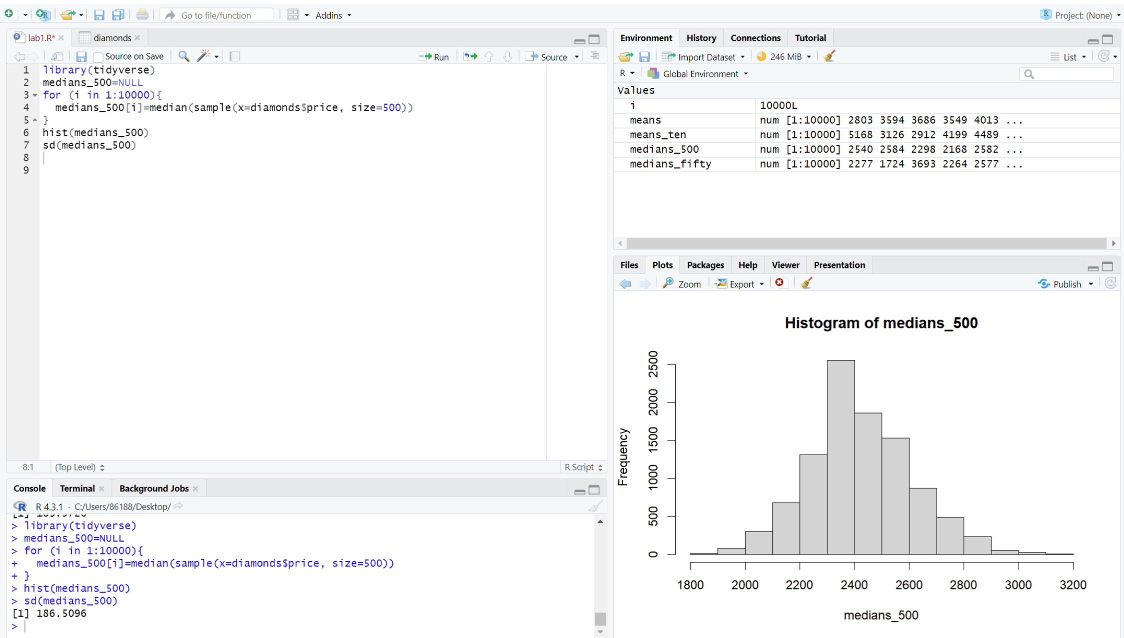
2. shape

   Right skewed. Precision: With a much larger sample size, the histogram may become normally distributed.

3. standard_deviation_6 = 618.7487

Q7.

1. histogram & code



2. shape:

   Right skewed. Compared to the "Histogram of medians_fifty", the "Histogram of medians_500" is less skewed, and closer to normal distribution.

3. standard_deviation_7 = 186.5096, less than the SD of medians_fifty. Expected.