Homework1

Xueying Liu

8/28/2020

Problem 2

Part A

```
library(knitr)
```

```
## Warning: package 'knitr' was built under R version 3.6.3
```

```
knitr::kable(c("R","Python","SAS"),col.names = "3 Things I Want to Learn")
```

3 Things I Want to Learn
R
Python
SAS

Part B

Bernoulli distribution(p):

$$P(X = x|p) = p^x(1-p)^{1-x};\ \ x = 0, 1;\ 0 <= p <= 1.$$

Binomial distribution(n,p):

$$P(X = x|n, p) = \binom{n}{x} p^x(1-p)^{n-x};\ \ x = 0, 1, 2, ..., n;\ 0 <= p <= 1.$$

Poisson distribution(λ):

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!};\ \ x = 0, 1, ...;\ 0 <= \lambda < \infty.$$

Problem 3

Steps in performing reproducible research:

*Step*1: For every result, keep track of how it was produced, record every involved steps. Challenges:Sometimes we just try some possible methods, manually recording each step can take a lot of works.

*Step*2: Archive the exact versions of all external programs used. Challenges:Remember to note the exact names and versions of the main programs you use.

*Step*3: Use version control system to track evolution of codes and help the exact reproduction of results.

*Step*4: Record all intermediate result, randomness and raw data behinds plots. Connect textual statements to underlying results. Challenges: It may be hard to locate the exact result underlying and supporting the statement from a large pool of different analyses with various versions.

*Step*5: Provide public access to your codes and results. Challenges: Be prepared to respond to any requests for further data or methodology details by peers.

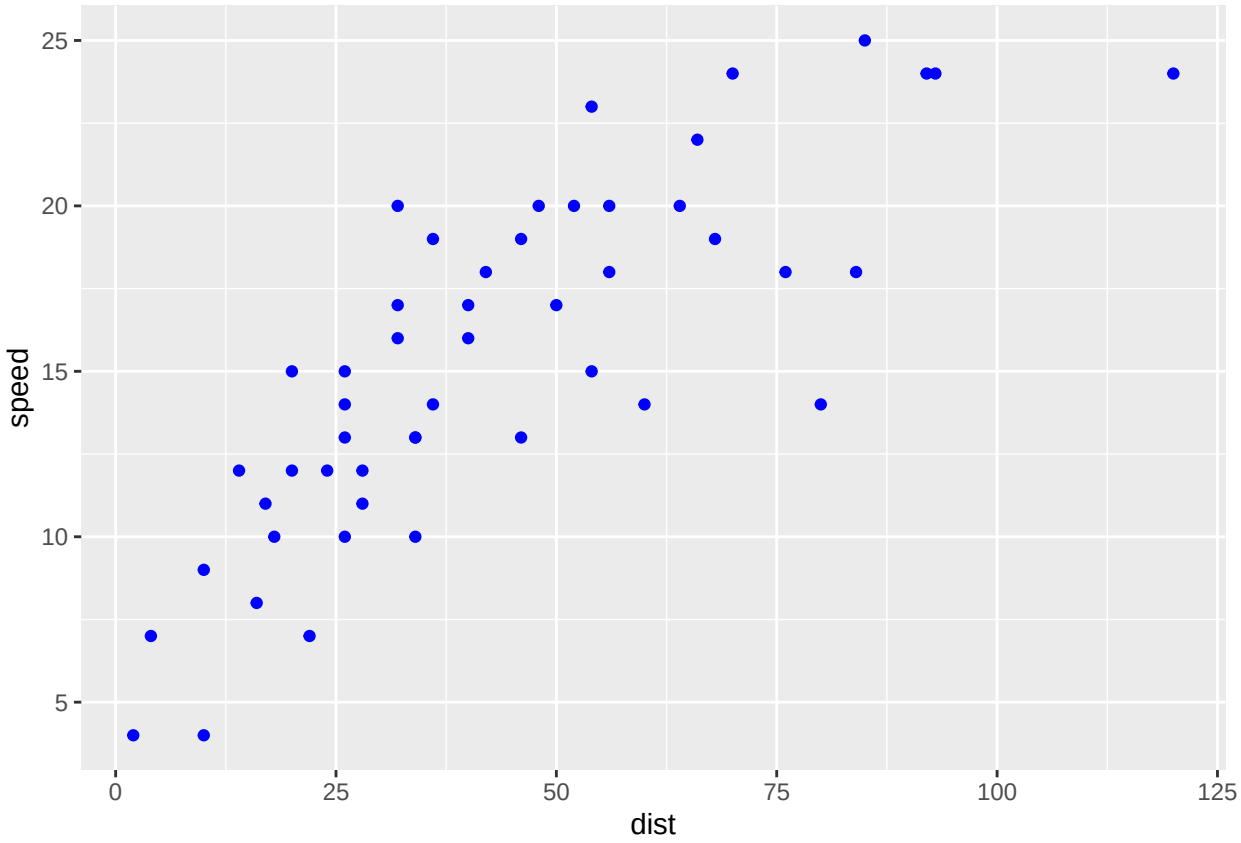
Problem 4

We will use the cars dataset in the R and plot a scatter plot and a histogram of it.

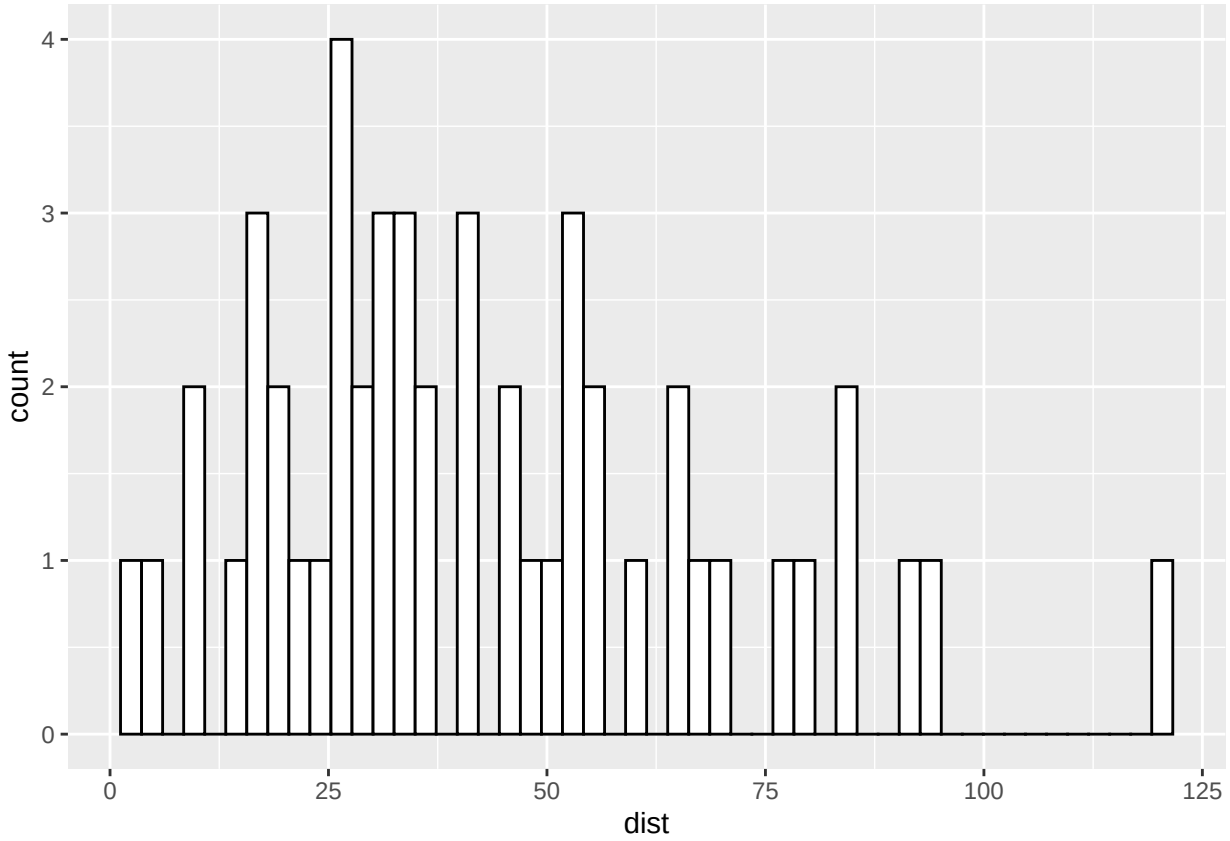
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

```
ggplot(data = cars,mapping = aes(x=dist, y=speed))+
  geom_point(color="blue")
```




```
ggplot(data=cars,aes(x=dist))+
  geom_histogram(color="black",fill="white", bins=50)
```


count
dist
0
1
2
3
4
0
25
50
75
100
125