

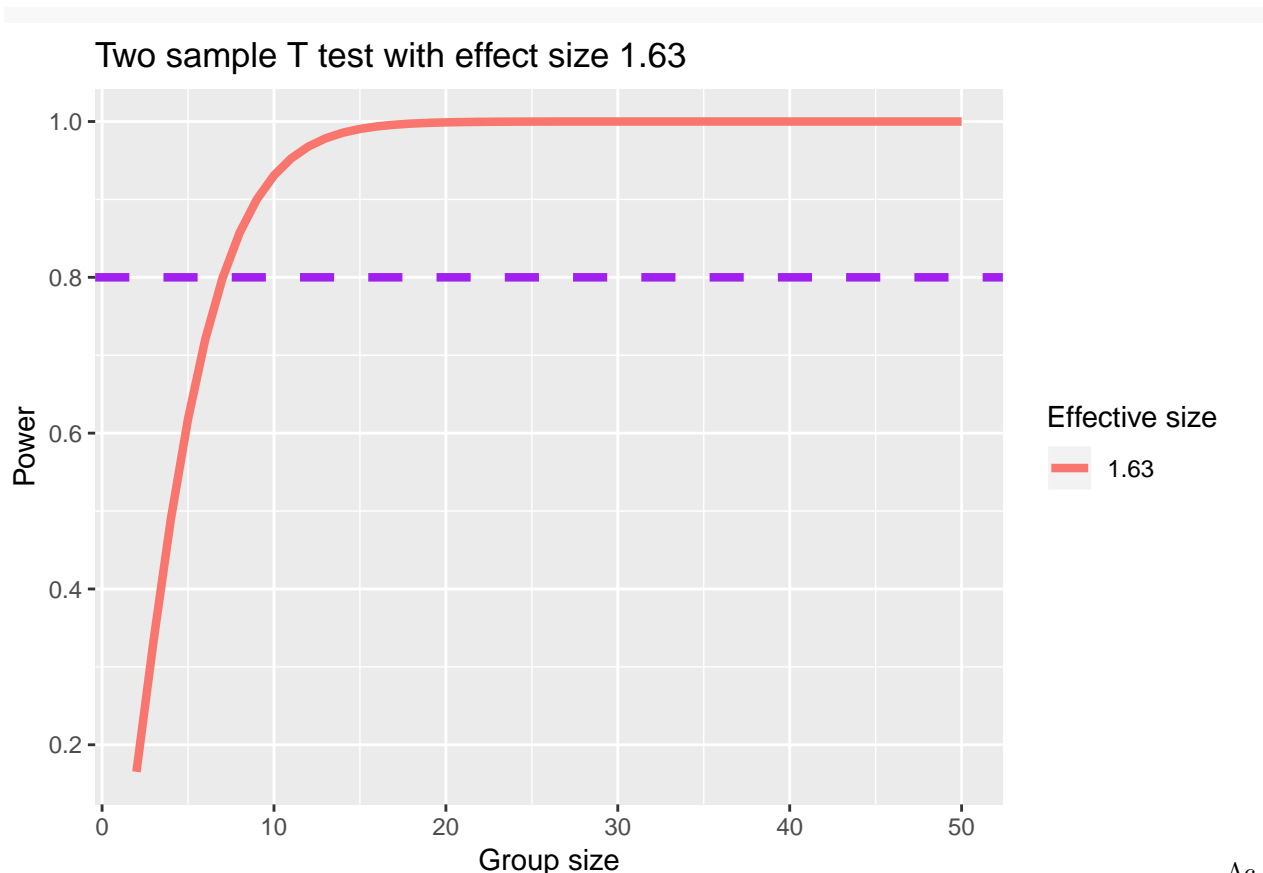
Jiahao Xu

H0: mortgage refusal rate of white applicants are the same as that of minority applicant H1: mortgage refusal rate of white applicants are lower than that of minority applicant

```
## [1] 1.279668e-07
```

```
# Power analysis to show the sufficiency
# Calculate the effect size
effect_size=abs(mean(Min)-mean(White))/sd(Min)

ptab1 <- cbind(NULL)
n <- seq(2, 50, by = 1)
for (i in seq(2, 50, by = 1)) {
  pwrt1 <- pwr.t2n.test(
    n1 = i, n2 = i,
    sig.level = 0.05, power = NULL,
    d = effect_size, alternative = "two.sided"
  )
  ptab1 <- rbind(ptab1, pwrt1$power)
}
temp<-as.data.frame(ptab1)
colnames(temp)[1]<-"num"
ggplot(temp) +
  geom_line(aes(x = n, y = num, colour = "darkblue"), size = 1.5) +
  scale_color_discrete(name = "Effective size", labels = c(round(effect_size,2))) +
  geom_hline(yintercept = 0.8, linetype = "dashed", color = "purple", size = 1.5) +
  ylab("Power") + scale_y_continuous(breaks = seq(0, 1, by = 0.2)) + ggtitle("Two sample T test with ef")
```



According to this power analysis plot, if we want to reach the general acceptable power 0.8, we need to have more than 7 data samples in each group under the effect size 1.63. And in our acorn data, we have 20 data samples in each group. Therefore, the data has sufficient evidence to show the discrimination to warrant corrective action.

2 Comparing Suppliers Revenue aside, which of the three schools produces the higher quality ornithopters, or are do they all produce about the same quality?

H0: They all produce about the same quality H1: They do not produce about the same quality

```
data2 <- matrix(c(12,23,89,8,12,62,21,30,119),ncol=3,nrow = 3,byrow=TRUE)
colnames(data2) <- c("dead","art","fly")
rownames(data2) <- c("Area51","BDV","Giffen")
fly <- as.table(data2)
chisq.test(data2,correct = F)
```

```
##
## Pearson's Chi-squared test
##
## data: data2
## X-squared = 1.3006, df = 4, p-value = 0.8613
```

The p-value of this chi-square test is 0.8613, which is much greater than the significant level $\alpha=0.05$. Therefore, we fail to reject the null hypothesis. The data are sufficient to show that three schools produce

the same quality.

3 How deadly are sharks?

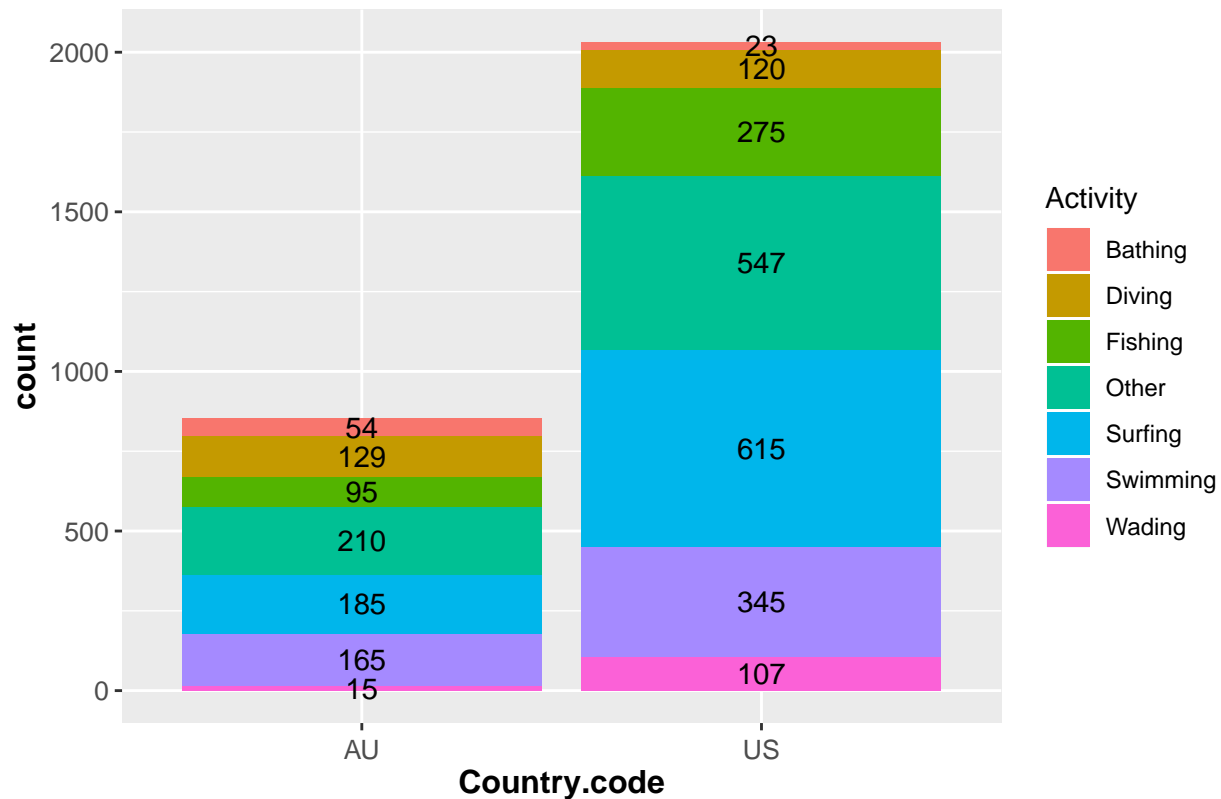
H0:Sharks in Australia were, on average, are the same as the sharks in the United States H1:Sharks in Australia were, on average, are more vicious lot than the sharks in the United States

If we want to analyze the difference of viciousness between US sharks and AU sharks, we only care about the attacks in AU and US and the sharks are not be provoked first.

```
data3<-read.csv("sharkattack.csv")
data3<-data3%>%filter(Country.code=="US"|Country.code=="AU"&Type=="Unprovoked")
temp<-data3%>%group_by(Country.code,Activity)%>%summarise(count=n())%>% ungroup()%>%group_by(Country.code)
kable(temp)
```

| Country.code | Activity | count | percent |
|--------------|----------|-------|-----------|
| AU | Bathing | 54 | 0.0633060 |
| AU | Diving | 129 | 0.1512309 |
| AU | Fishing | 95 | 0.1113716 |
| AU | Other | 210 | 0.2461899 |
| AU | Surfing | 185 | 0.2168816 |
| AU | Swimming | 165 | 0.1934349 |
| AU | Wading | 15 | 0.0175850 |
| US | Bathing | 23 | 0.0113189 |
| US | Diving | 120 | 0.0590551 |
| US | Fishing | 275 | 0.1353346 |
| US | Other | 547 | 0.2691929 |
| US | Surfing | 615 | 0.3026575 |
| US | Swimming | 345 | 0.1697835 |
| US | Wading | 107 | 0.0526575 |

```
ggplot(data = temp, aes(Country.code, count, group = Activity)) +
  geom_col(aes(fill = Activity)) +
  geom_text(aes(label = count), position = position_stack(vjust = 0.5))+theme(axis.title.x = element_text(
    axis.title.y = element_text(face="bold", size=12),
    plot.title = element_text(size=12, face="bold"),
    axis.text.x = element_text(vjust=0.5, size=10),axis.text.y = element_text(vjust=0.5, size=10)
```



According to the bar plot and data frame, US sharks made more attacks in total and attacks in surfing had higher percentage than AU sharks.

```
# Transfer dataframe into matrix to do chi-square test
data33<- matrix(c(23,120,275,547,615,347,107,54,129,95,210,186,165,12), nrow=2,
               dimnames = list(c("AU","US"),c("Bathing","Diving","Fishing","Other","Surfing","Swimming","Wading")),
               chisq.test(data33,correct = F))
```

```
##
## Pearson's Chi-squared test
##
## data: data33
## X-squared = 378.78, df = 6, p-value < 2.2e-16
```

The p-value of this chi-square test is much smaller than the significant level $\alpha=0.05$. Therefore, we reject the null hypothesis. Sharks in Australia were, on average, are more vicious than the sharks in the United States. The statistical power of the test chi-square test is 1.

4 Power analysis

The arcsine transformation is calculated as two times the arcsine of the square root of the proportion. Doing this can transform the proportional parameter from (0,1) to $(-\pi/2, \pi/2)$. The effect of the arcsine transformation is similar to the logit, in that it pulls out the ends of the distribution, but not to the extent that the logit does.

The power to detect the hypothetical parameters .65 and .45 is .48 while the power to detect the difference between hypothetical parameters .25 and .05 is .82, even though the difference between both pairs of values is .20. The reason why this happens is that 0.25 and 0.05 are at the end (extreme value) of the distribution. Therefore, it takes more power to detect and arcsine transformation can solve this problem.

5 Use the Method of Moments and MLE to find estimators as described in these three cases.

Case1 MLE of Exponential Distribution

$$f(x; \lambda) = \lambda e^{-\lambda x}$$

$$L(\lambda; X_1, \dots, X_n) = \lambda e^{-\lambda X_1} \lambda e^{-\lambda X_2} \dots \lambda e^{-\lambda X_n}$$

$$L(\lambda; X_1, \dots, X_n) = \lambda^n e^{-\lambda \sum X_i}$$

$$l(\lambda; X_1, \dots, X_n) = n \log(\lambda) - \lambda \sum X_i$$

$$\frac{dl(\lambda; X_1, \dots, X_n)}{d\lambda} = \frac{n}{\lambda} - \sum X_i = 0$$

$$\hat{\lambda} = \frac{n}{\sum X_i} = \frac{1}{\bar{X}_n}$$

Case2 Moment Estimator and MLE for new distribution θ

MOM:

$$\begin{aligned} E[X] &= \int_0^1 x((1-\theta) + 2\theta x) dx \\ &= (1-\theta) \int_0^1 x dx + \int_0^1 2\theta x^2 dx \\ &= (1-\theta) \frac{1}{2} x^2 \Big|_0^1 + 2\theta \frac{1}{3} x^3 \Big|_0^1 \\ &= \frac{1}{2} - \frac{1}{2}\theta + \frac{2}{3}\theta \\ &= \frac{1}{6}\theta + \frac{1}{2} \end{aligned}$$

MLE:

$$L(\theta; X_1, \dots, X_n) = [(1-\theta) + 2\theta X_1] \dots [(1-\theta) + 2\theta X_n]$$

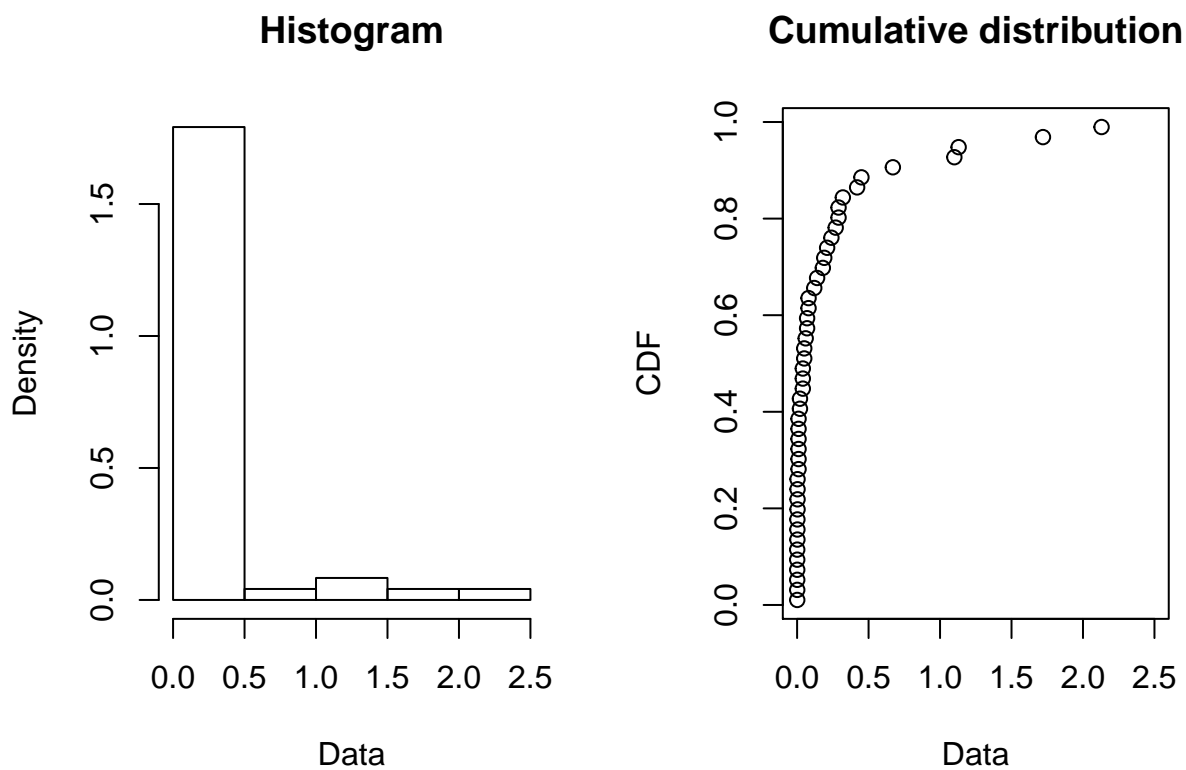
$$l(\theta; X_1, \dots, X_n) = \log[(1-\theta) + 2\theta X_1] + \dots + \log[(1-\theta) + 2\theta X_n]$$

Under this situation, we cannot taking the derivative of $l(\theta; X_1, \dots, X_n)$ to find the maximum value with corresponding θ .

Case3 Rainfall analysis

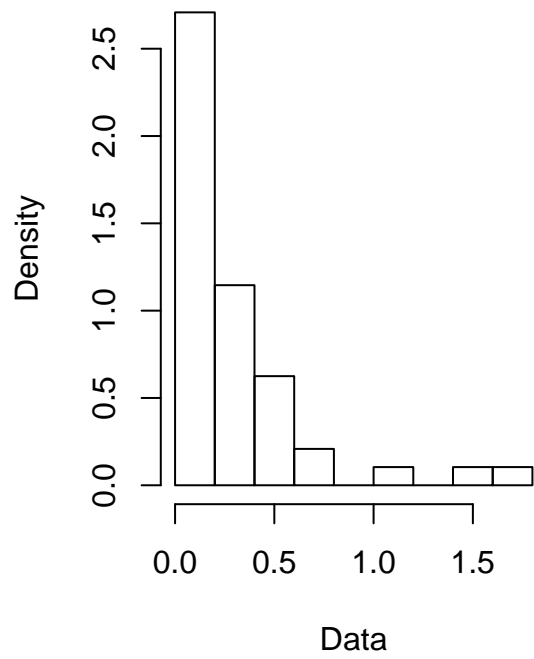
```
# read the data
data60 <- read.table("ill-60.txt", quote="\"", comment.char="")
data60 <- as.numeric(as.array(data60[,1]))
data61 <- read.table("ill-61.txt", quote="\"", comment.char="")
data61 <- as.numeric(as.array(data61[,1]))
data62 <- read.table("ill-62.txt", quote="\"", comment.char="")
data62 <- as.numeric(as.array(data62[,1]))
data63 <- read.table("ill-63.txt", quote="\"", comment.char="")
data63 <- as.numeric(as.array(data63[,1]))
data64 <- read.table("ill-64.txt", quote="\"", comment.char="")
data64 <- as.numeric(as.array(data64[,1]))

# explore the distribution of the rainfall data
plotdist(data60)
```

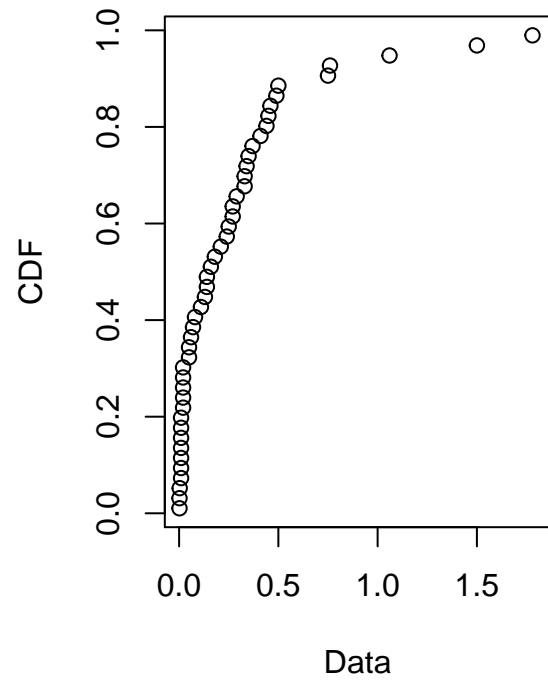


```
plotdist(data61)
```

Histogram

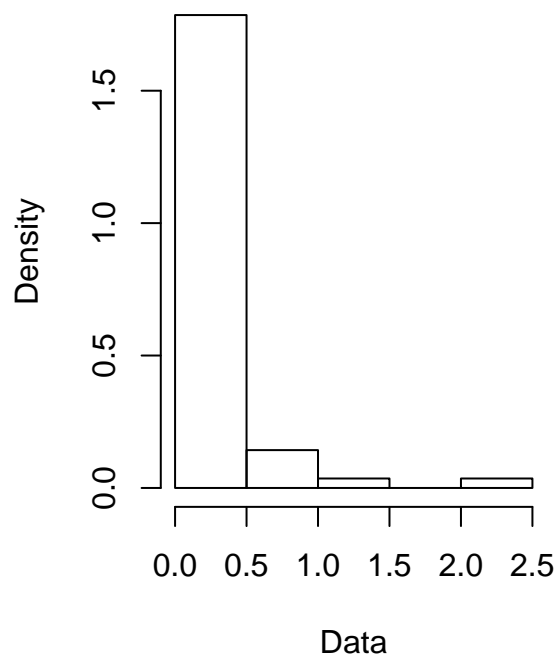


Cumulative distribution

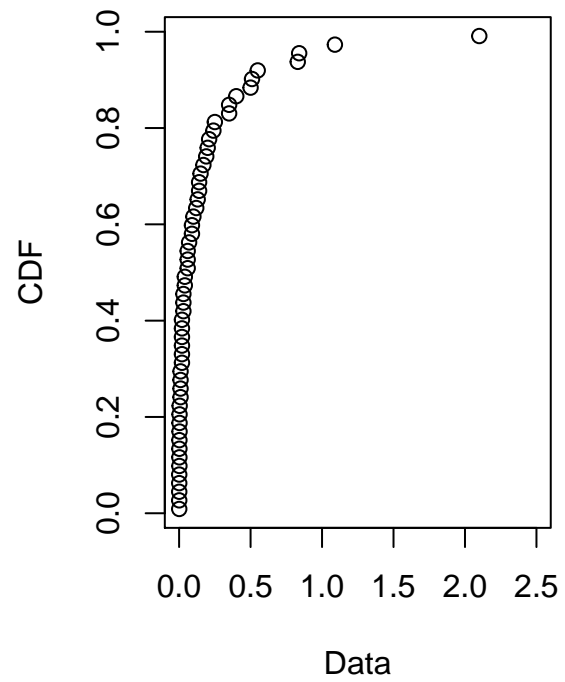


```
plotdist(data62)
```

Histogram

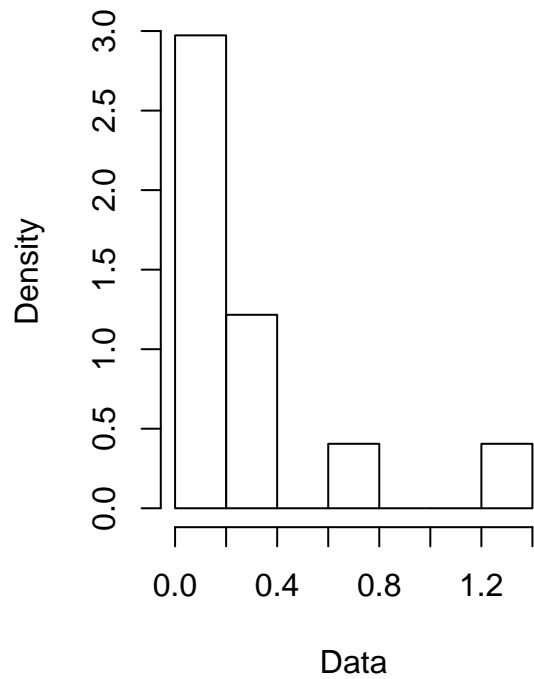


Cumulative distribution

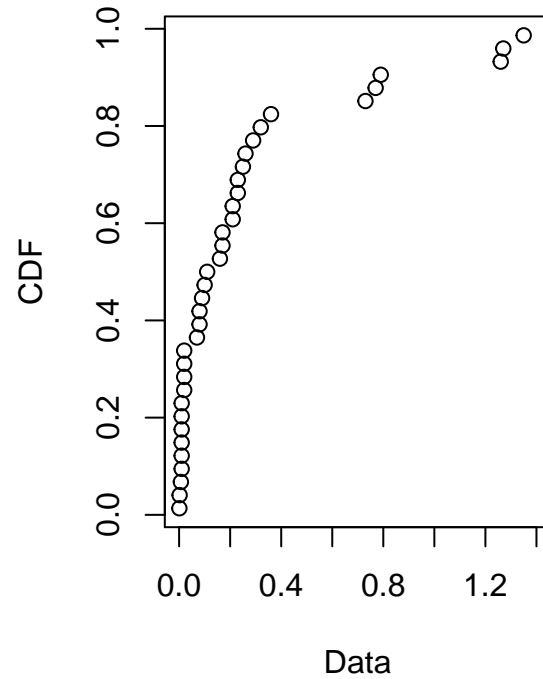


```
plotdist(data63)
```

Histogram

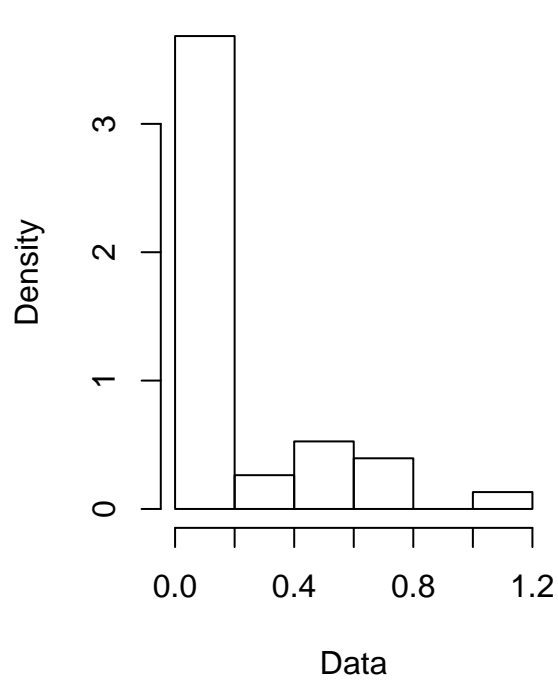


Cumulative distribution

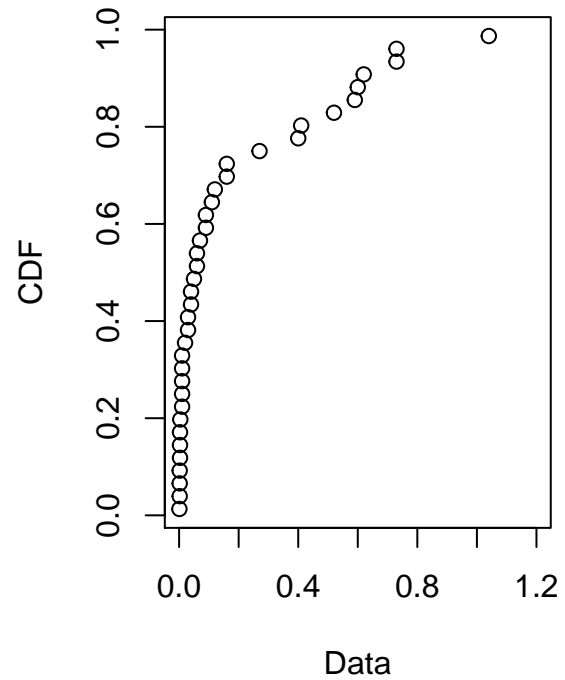


```
plotdist(data64)
```

Histogram



Cumulative distribution

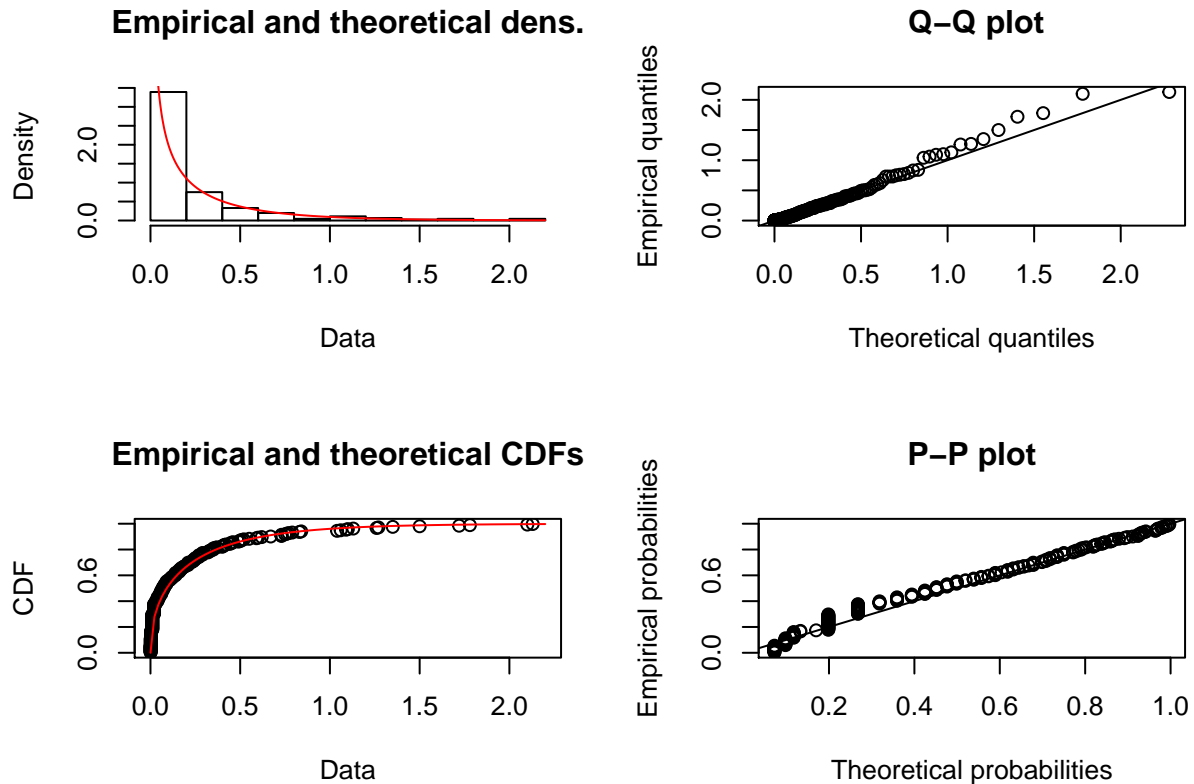


```
Rain<-matrix(c(sum(data60),sum(data60)/length(data60),sum(data61),sum(data61)/length(data61),sum(data62),sum(data62)/length(data62)),nrow=100,ncol=5)
kable(Rain)
```


| | 1960 | 1961 | 1962 | 1963 | 1964 |
|-------------|------------|------------|----------|-----------|-----------|
| TotalRain | 10.5740000 | 13.1970000 | 10.34600 | 9.7100000 | 7.1100000 |
| AverageRain | 0.2202917 | 0.2749375 | 0.18475 | 0.2624324 | 0.1871053 |

According to the distribution plot, five years are similar. 1961 is more wetter than others since it has the highest total rainfall. And from the row AverageRain in table, we can see that 1961 and 1963 have similar average rainfall in each storm. Therefore, it should be that storms produced more rain.

```
#Test whether the gamma distribution was a good fit for their data.
alldata<-c(data60,data61,data62,data63,data64)
test_gamma <- fitdist(alldata, "gamma")
plot(test_gamma)
```



According to the Empirical the theoretical CDFs and P-P plot, the data points almost excellent fit the gamma distribution. Although in Q-Q plot, the points after 1.0 theoretical quantiles are a little bit discrete from the gamma distribution. Generally, the gamma distribution was a good fit for their data. I will agree with Changnon and Hu.

```
# calculate MOM and MLE
set.seed(2019)
mom <- fitdist(alldata, "gamma", method = "mme")
boot_mom <- bootdist(mom)
summary(boot_mom)

## Parametric bootstrap medians and 95% percentile CI
##           Median      2.5%      97.5%
## shape 0.3902039 0.2777425 0.5301561
## rate  1.7497653 1.1801093 2.5899000

mle <- fitdist(alldata, "gamma", method = "mle")
boot_mle <- bootdist(mle)
summary(boot_mle)
```

```
## Parametric bootstrap medians and 95% percentile CI
##           Median      2.5%      97.5%
## shape 0.4447087 0.3808683 0.5170126
## rate  1.9913553 1.5616429 2.5430885
```

According to summary of MOM, the 95% confidence interval of shape from bootstrap sample is (0.27,0.53), the rate is (1.18,2.59). For MLE, the 95% confidence interval of shape from bootstrap sample is (0.38,0.51),the rate is (1.56,2.54).The MLE estimates have higher rate and more narraow CI, which means lower variance.Therefore, I would choose to present MLE as the estimator.

Analysis of decision theory article

#Derive equations (10a), (10b), (10c) in Section 3.2.2.

#Use R to reproduce the calculations in Table 1 which is explained in 3.2.3. Describe what you have don