Beyond Capes and Cloaks

Casey Hutchinson

2023-11-01

Iconix Talent Agency - Insights

Questions:

1. Construct histograms of Intellect, Strength, Adaptability, Charisma, and Aggregate and use them to answer the following:

a. Between Intellect, Strength, and Aggregate, classify each as symmetric, skewed to the left, or skewed to the right.
b. Between Intellect and Adaptability, classify each as unimodal or bimodal.
c. Of all the variables, which appears to be approximately uniform?

2. A secret document reports that Strength and Adaptability are both normally distributed. Construct a stripchart and normal quantile plot for each and explain whether these plots support the claims.

3. Verify that the variable Aggregate is approximately normal. Use the Empirical Rule to complete the following statements:

a. Approximately 68% of the superheroes have Aggregate scores between _____ and _____.
b. Approximately 95% of the superheroes have Aggregate scores between _____ and _____.

4. Use the pnorm() command to verify that the statements in #3 are correct.

5. Iconix is considering dropping all superheroes whose Aggregate score is below 99. Find the percentage of superheroes that would be dropped from this list.

6. A superhero is randomly chosen. What is the probability that their Aggregate score is greater than 400?

7. Use the command filter(… , Alignment == “good”) to create a dataset comprised of good superheroes. Use a similar command to create a dataset of bad superheroes. On average, which group scores higher on intellect?

8. Iron Man has a Strength score of 84.13 and is a good superhero. Find the z-score of Iron Man’s strength. Ultron also has a Strength score of 84.13 and is a badsuperhero. Find the z-score of Ultron’s strength. Which is stronger relative to their ethical alignment?

Data

The data in this study can be found here: Iconix Talent Agency Data (http://datasets.barrymonk.com/MATH3440/Iconix_Superhero_Talent_Agency.csv)

Dataset Names

```
library(readr)
Superhero <- read.csv("http://datasets.barrymonk.com/MATH3440/Iconix_Superhero_Talent_Agency.csv")
```

I will also add the dataset names *Good.Superhero* and *Bad.Superhero* when we dive deeper into the data.

Question 1) Construct histograms of *Intellect*, *Strength*, *Adaptability*, *Charisma*, and *Aggregate* and use them to answer the following questions.

a) Between Intellect, Strength, and Aggregate, classify each as symmetric, skewed to the left, or skewed to the right.

The histogram for Intellect is skewed to the left, the histogram for Strength is skewed slightly to the right, and the histogram for Aggregate is symmetric.

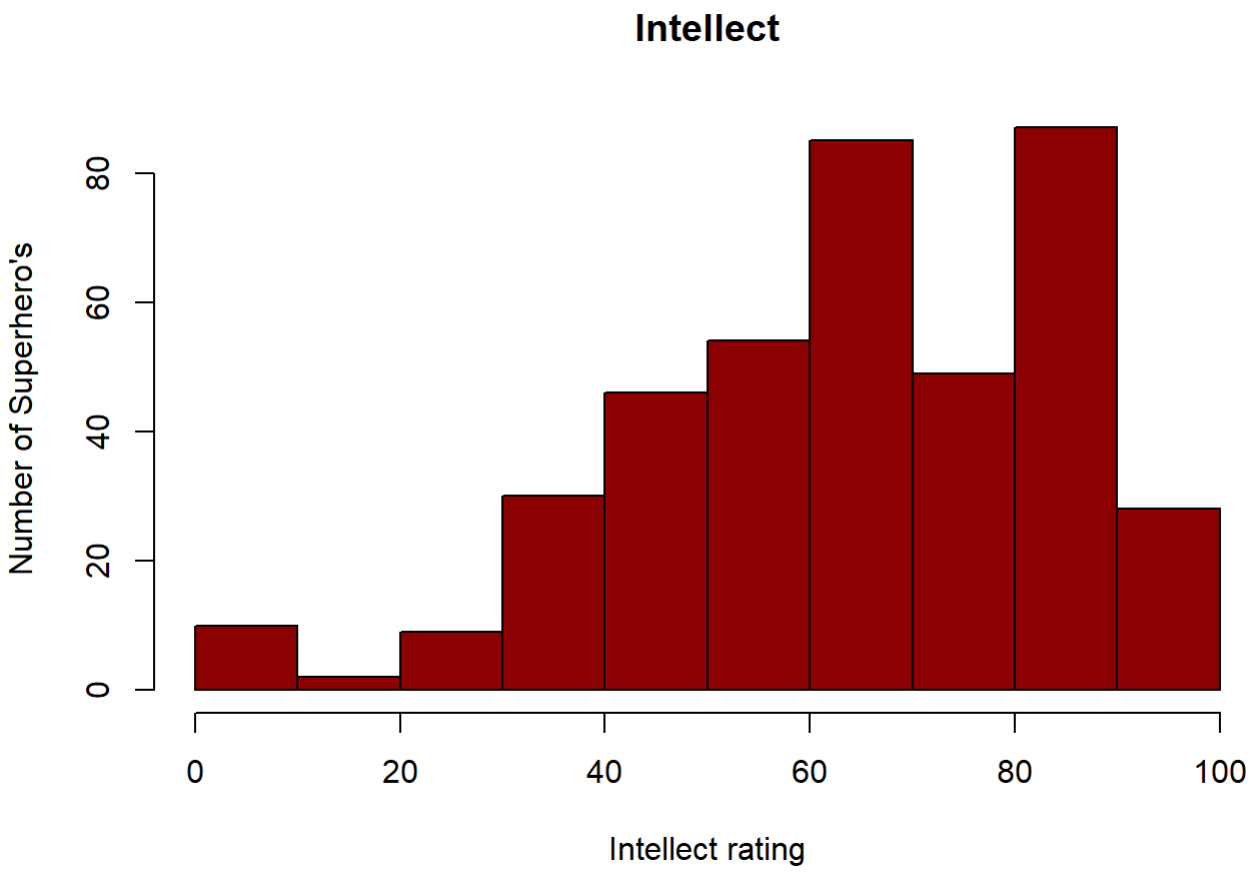
b) Between Intellect and Adaptability, classify each as unimodal or bimodal.

Examining the Intellect histogram you can see there are 2 distinct peaks, therefore we'd classify the Intellect graph as bimodal. When you compare it to the histogram for Adaptability, you can see that it only has one peak and is therefore unimodal.

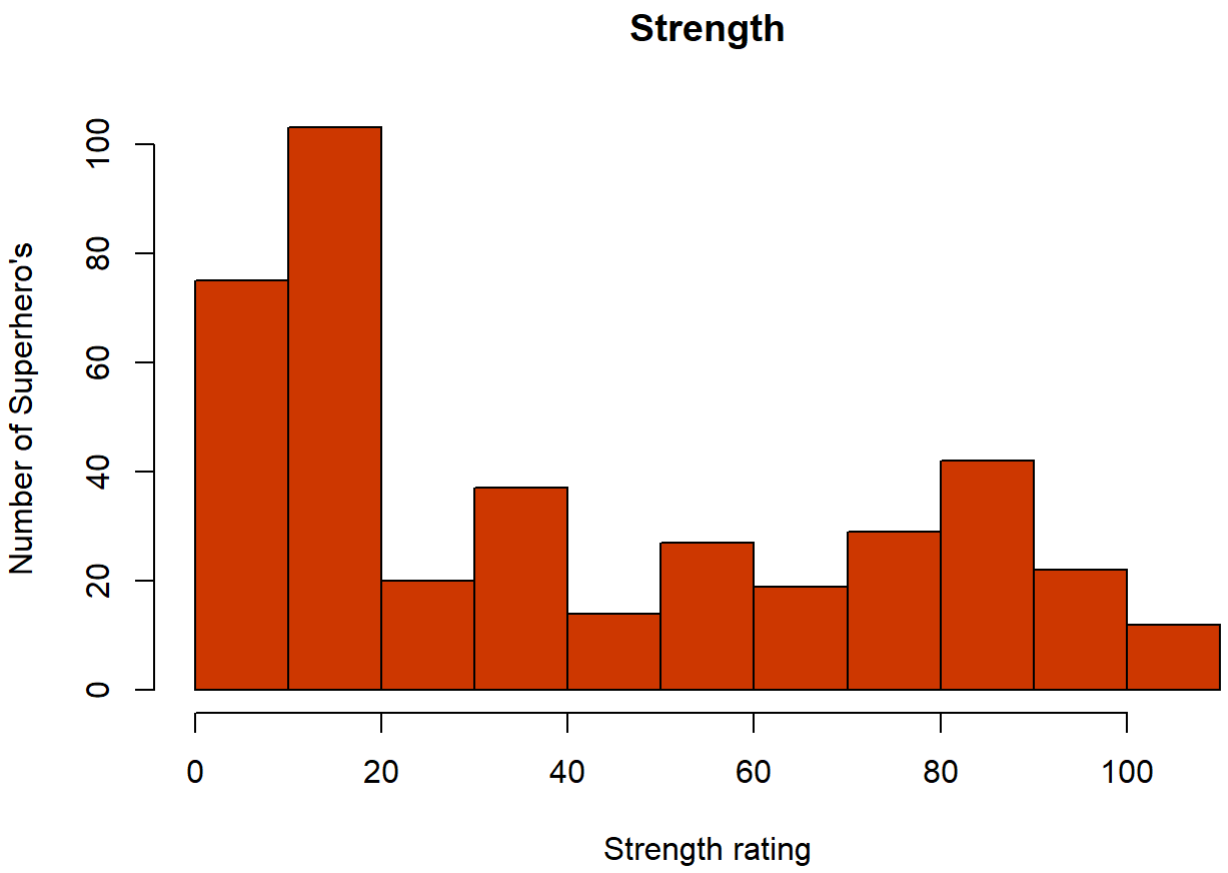
c) Of all the variables, which appears to be approximately uniform?

Upon examining the Charisma histogram, you can see that it is fairly uniform.

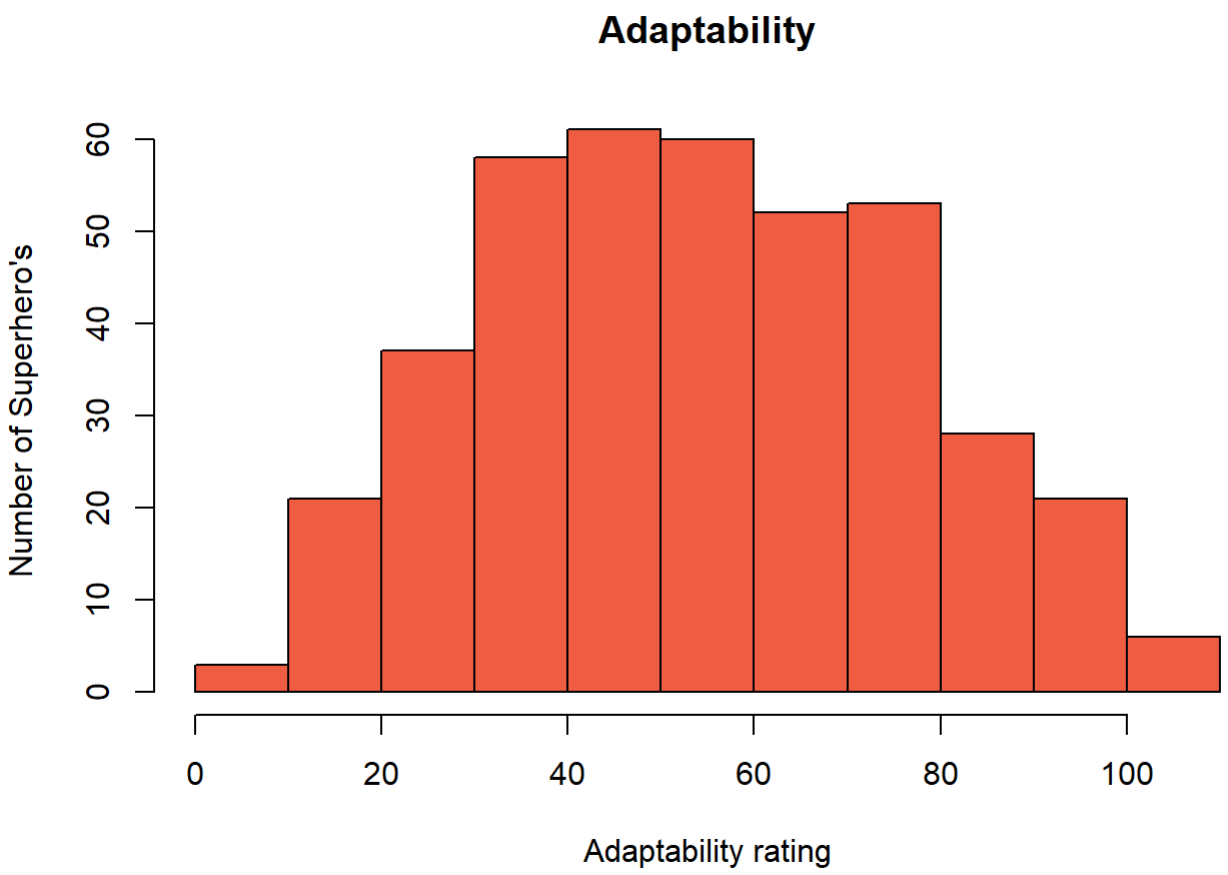
```
hist(Superhero$Intellect, main="Intellect", xlab="Intellect rating", ylab="Number of Superhero's", col="red4", breaks = 10)
```



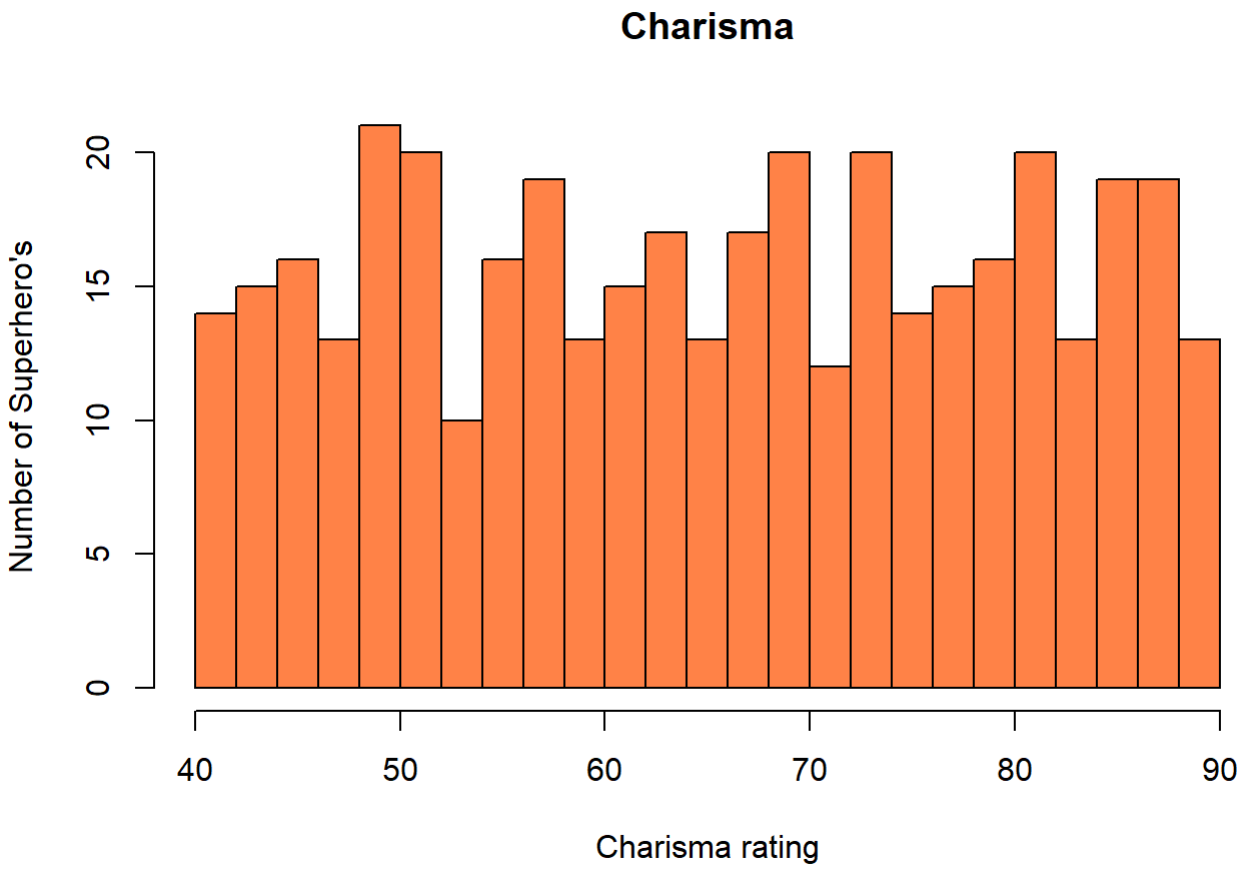

```
hist(Superhero$Strength, main="Strength", xlab="Strength rating", ylab="Number of Superhero's",
col="orangered3", breaks = 11)
```



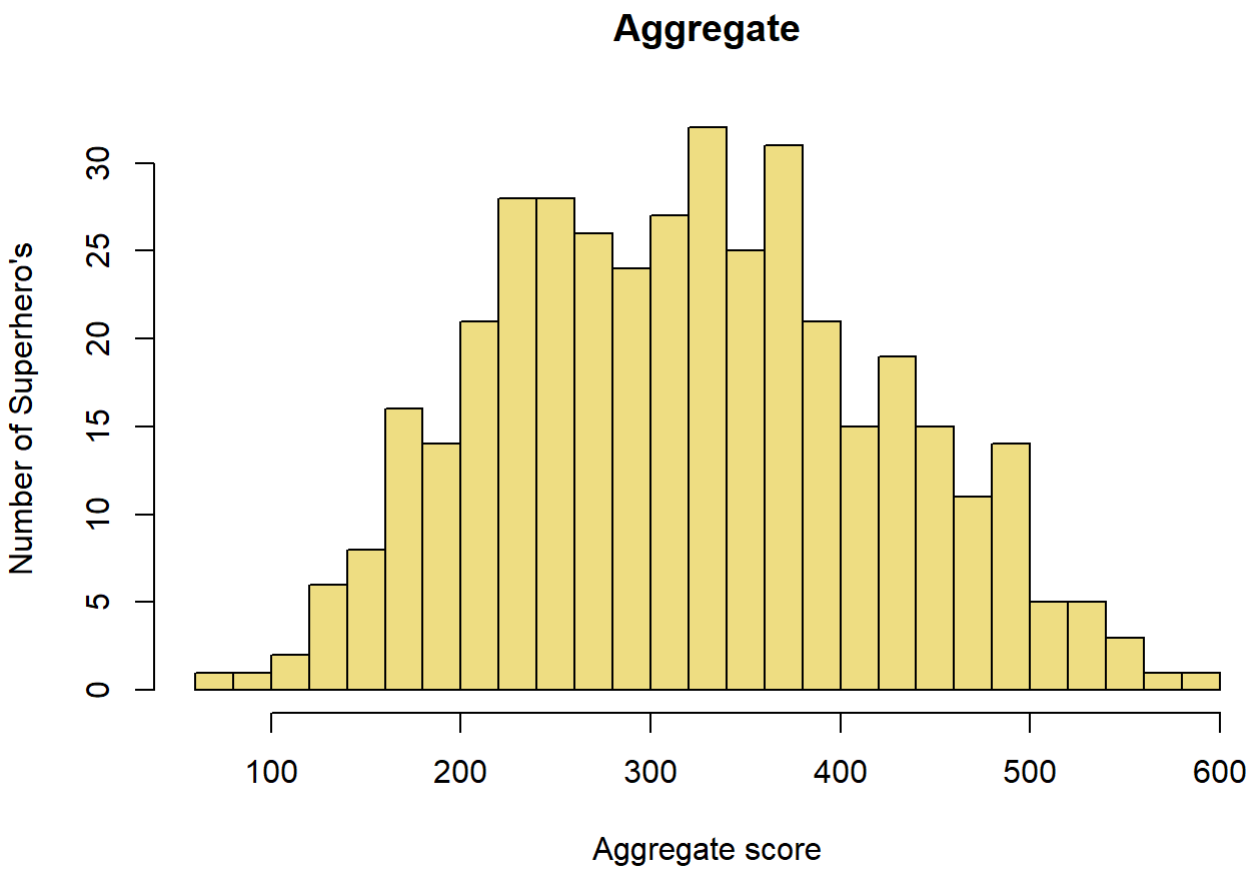

```
hist(Superhero$Adaptability, main="Adaptability", xlab="Adaptability rating", ylab="Number of Su
perhero's", col="tomato2", breaks = 11)
```




```
hist(Superhero$Charisma, main="Charisma", xlab="Charisma rating", ylab="Number of Superhero's",
col="sienna1", breaks = 25)
```

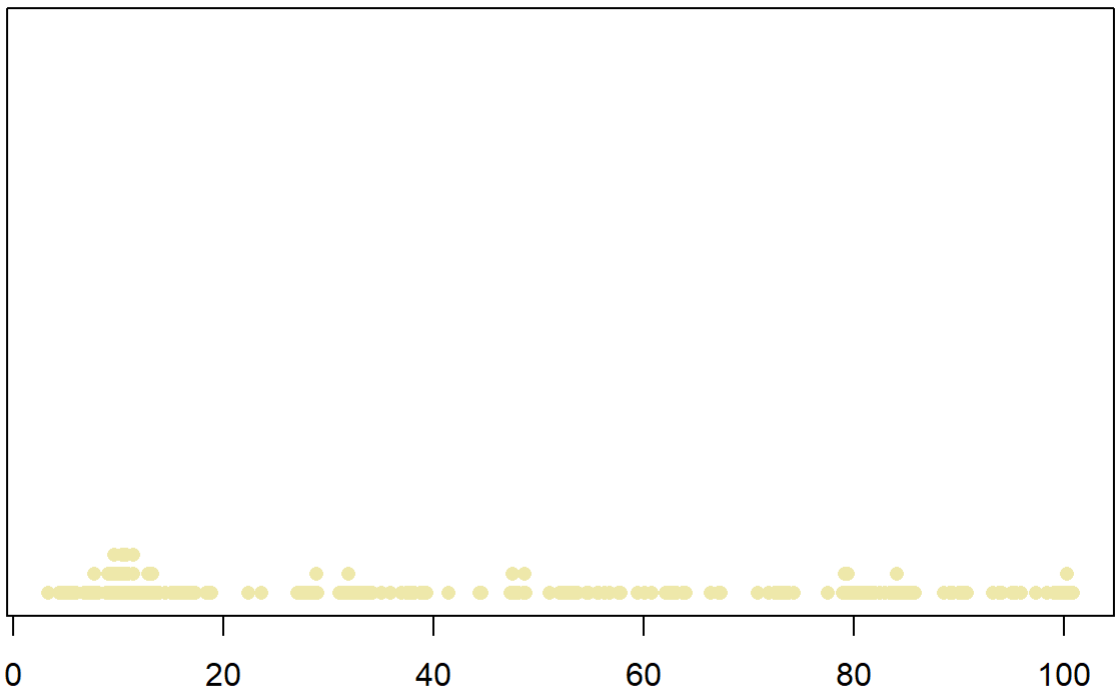



```
hist(Superhero$Aggregate, main="Aggregate", xlab="Aggregate score", ylab="Number of Superher
o's", col="lightgoldenrod", breaks = 25)
```

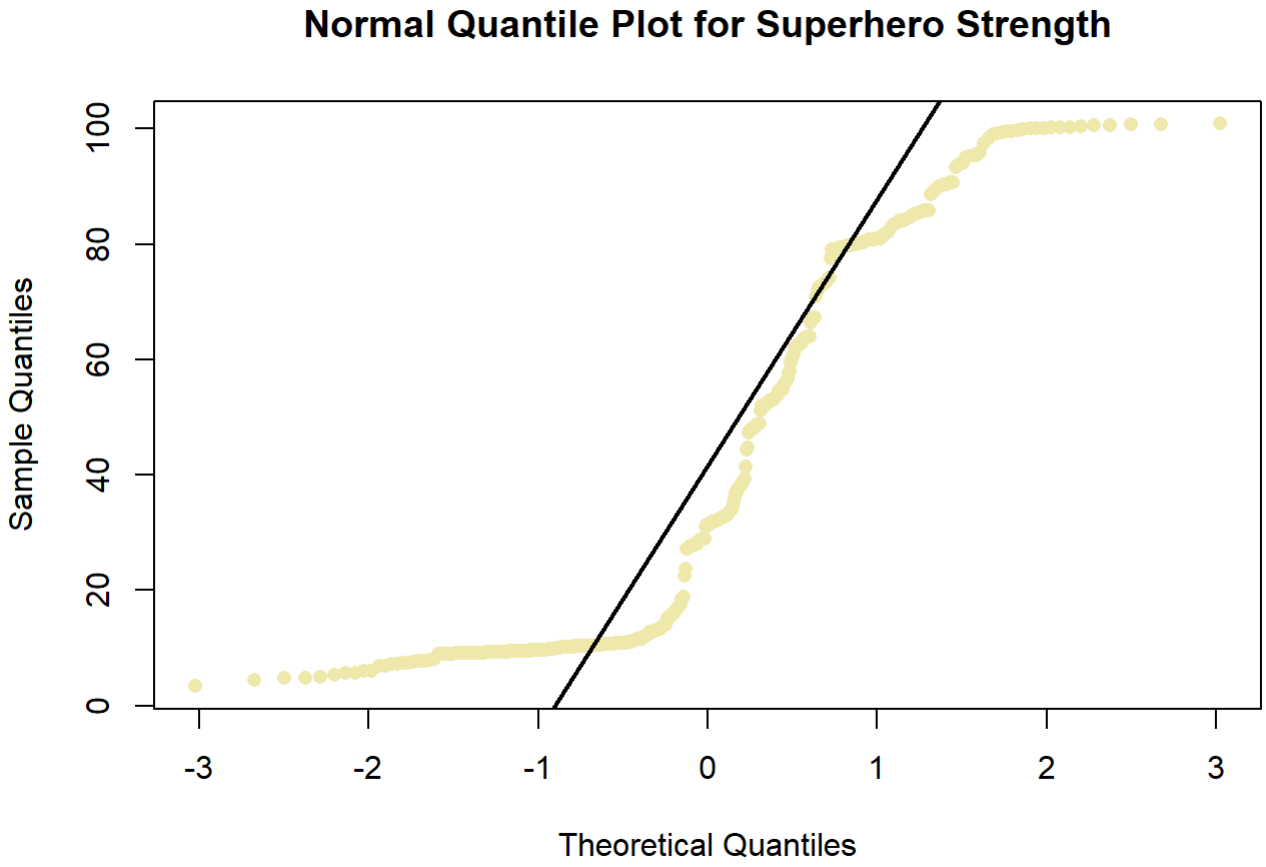


Question 2) A secret document reports that *Strength* and *Adaptability* are both normally distributed. Construct a stripchart and normal quantile plot for each and explain whether these plots support the claims.

```
stripchart(Superhero$Strength, main="Stripchart for Superhero Strength", method="stack", at=0, p
ch=16, col="palegoldenrod", offset=0.5)
```

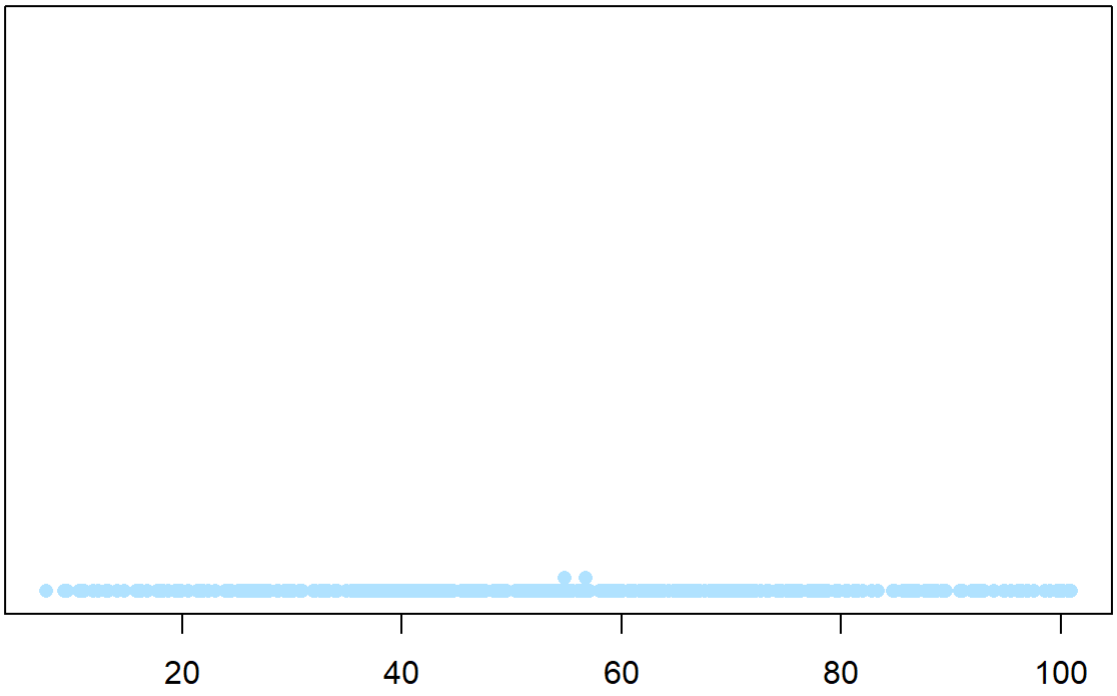



```
qqnorm(Superhero$Strength, main="Normal Quantile Plot for Superhero Strength", pch=16, col="pale goldenrod")
qqline(Superhero$Strength, lwd=2)
```

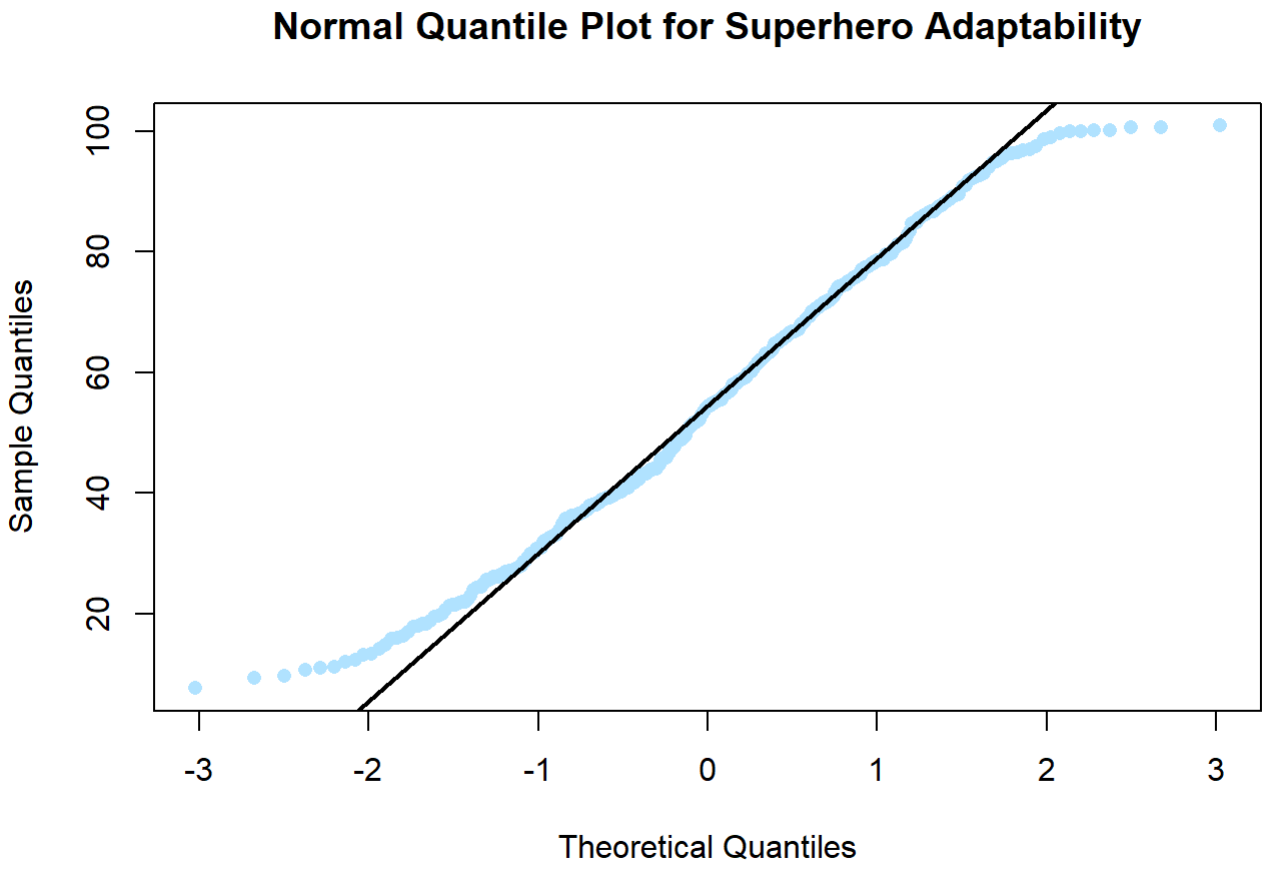


A first glance at the stripchart doesn't suggest anything too out of the ordinary, but we can see on the left-hand side a concentration that may suggest multiple modes. Graphing out a normal quantile plot shows points differing heavily from the line - mostly to the left, but also some to the right. There is an argument to be made that it looks like the bulk of the data points are within 1 theoretical quantiles of the line, but it skews too heavily in my opinion to say that Strength is distributed normally. So these plots do not support the suggestion that Strength is normally distributed.

```
stripchart(Superhero$Adaptability, main="Stripchart for Superhero Adaptability", method="stack",
at=0, pch=16, col="lightskyblue1")
```




```
qqnorm(Superhero$Adaptability, main="Normal Quantile Plot for Superhero Adaptability", pch=16, c
ol="lightskyblue1")
qqline(Superhero$Adaptability, lwd=2)
```



The stripchart for Adaptability does not have the same concentration at any point on its graph that Strength had. In fact, it looks fairly evenly distributed. The normal quantile plot supports this. There are some points on the left and right that differ heavily from the line, but they don't do so as heavily as they do in the Strength plot. The plots for Adaptability do support the suggestion that it is normally distributed.

Question 3) Verify that the variable *Aggregate* is approximately normal. Use the Empirical Rule to complete the following statements:

a) Approximately 68% of the superheroes have Aggregate scores between what 2 values?

b) Approximately 95% of the superheroes have Aggregate scores between what 2 values?

To find these values, we first need to know the Mean and Standard Deviation. We need these because the Emprical Rule states that approximately 68% of a normal distribution will be within one Standard Deviation of the Mean, and 95% will be within two Standard Deviations of the Mean.

```
M<-mean(Superhero$Aggregate); M
```

```
## [1] 320.3356
```

```
S<-sd(Superhero$Aggregate); S
```

```
## [1] 101.3333
```

```
M-S
```

```
## [1] 219.0024
```

```
M+S
```

```
## [1] 421.6689
```

One Standard Deviation to the left of the Mean is represented by M-S, and one Standard Deviation to the right of the Mean is represented by M+S. **We can see from the calculations that 68% of the Superheroes have Aggregate scores between 219 and 422.**

```
M-2*S
```

```
## [1] 117.6691
```

```
M+2*S
```

```
## [1] 523.0022
```

Two Standard Deviations to the left of the Mean is represented by M-2xS, and two Standard Deviations to the right of the Mean is represented by M+2xS. **We can see from the calculations that 95% of the Superheroes have Aggregate scores between 118 and 523.**

Question 4) Use the pnorm() command to verify that the statements in #3 are correct.

We'll test our values for 68% first:

```
pnorm(421.6689, M, S) - pnorm(219.0024, M, S)
```

```
## [1] 0.6826893
```

Using the values we calculated above with the pnorm command, **we see that we return the value 0.6827 - or approximately 68%.**

Now lets test our values for 95%:

```
pnorm(523.0022, M, S) - pnorm(117.6691, M, S)
```

```
## [1] 0.9544997
```

Using the values we calculated above with the pnorm command, **we see that we return the value 0.9545 - or approximately 95%.**

Question 5) Iconix is considering dropping all Superheroes whose *Aggregate* score is below 99. Find the percentage of Superheroes that would be dropped from the list.

```
pnorm(99, M, S)
```

```
## [1] 0.01447252
```

If Iconix were to drop all Superheroes whose Aggregate score is below 99, then they would end up dropping 1.4% from the list. I never thought I'd see the day that Kool-Aid Man wasn't marketable enough…

Question 6) A Superhero is randomly chosen. What is the probability that their *Aggregate* score is greater than 400?

```
1 - pnorm(400, M, S)
```

```
## [1] 0.2158864
```

If a Superhero were randomly chosen, the probability that their Aggregate score is greater than 400 is approximately 21.6%

Question 7) Use the command filter(…, Alignment == "good") to create a dataset comprised of good Superheroes. Use a similar command to create a dataset comprised of bad Superheroes. On average, which group scores higher on *Intellect*?

```
library('dplyr')
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
Good.Superhero <-filter(Superhero, Alignment == "good")
Bad.Superhero <-filter(Superhero, Alignment == "bad")
```

We'll start with finding the average Intellect in our Good Superhero dataset:

```
mean(Good.Superhero$Intellect)
```

```
## [1] 63.49112
```

And now we'll find the average Intellect in our Bad Superhero dataset:

```
mean(Bad.Superhero$Intellect)
```

```
## [1] 68.57588
```

Unfortunately, although perhaps unsurprising, we find that the average Intellect rating is higher for "Bad" Superheroes compared to "Good" Superheroes. It is close though, as the average rating for Bad is 68.58 and the average rating for Good is 63.49.

Question 8) Iron Man has a *Strength* score of 84.13 and is a good Superhero. Find the z-score of Iron Man's strength. Ultron has a *Strength* score of 84.13 and is a bad Superhero. Find the z-score of Ultron's strength.Which is stronger relative to their ehtical alignment?

We'll start with the Iron Man side of this comparison. To calculate the z-score of his strength, we'll need to find the Mean and Standard Deviation of strength in the Good.Superhero dataset first. Then, we can find his specific z-score.

```
MG <-mean(Good.Superhero$Strength)
SG <-sd(Good.Superhero$Strength)
(84.13-MG)/SG
```

```
## [1] 1.476074
```

Here we can see the z-score of Iron Man's strength comes out to **1.476074**. Now we will find the Mean and Standard Deviation of strength in the Bad.Superhero dataset so that we can find Ultron's z-score.

```
MB <-mean(Bad.Superhero$Strength)
SB <-sd(Bad.Superhero$Strength)
(84.13-MB)/SB
```

```
## [1] 1.160093
```

Here we can see the z-score of Ultron's strength comes out to **1.160093**.

It's only by a small amount, but relative to their ethical alignment, Iron Man is stronger than Ultron.