Normal distribution

Reflection

You know how to do the basics:

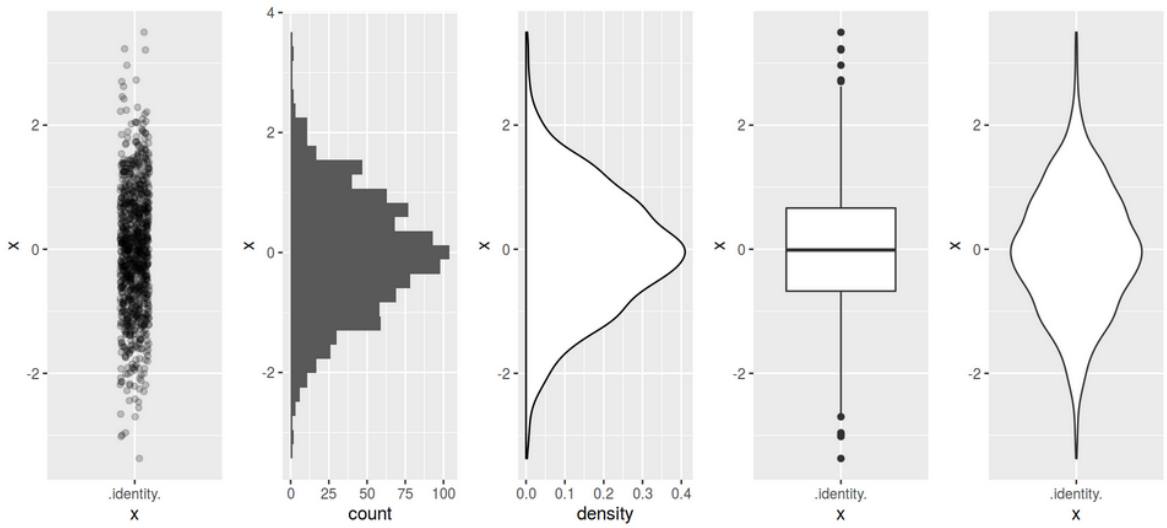
- read data into R,
- explore the data set,
- count some statistics,
- create and interpret basic plots,
- describe the plots with labels, change the style, save them.

Some additions…

- *Where do I get help?*
- In cheat sheets.
- *What type of graph should I choose?*
- Look in R Graph Gallery.
- *What colors should I use?*
- Look at Color Brewer.
- See section Resources at the website for more details…

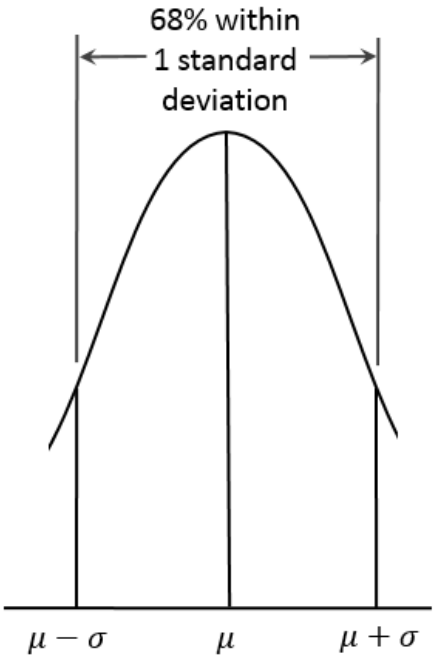
Normal distribution

bell-shaped curve, *Gaussian distribution*


x
2
0
-2
.identity.
x
4
2
0
-2
0
25
50
75
100
count
density
0.0
0.1
0.2
0.3
0.4

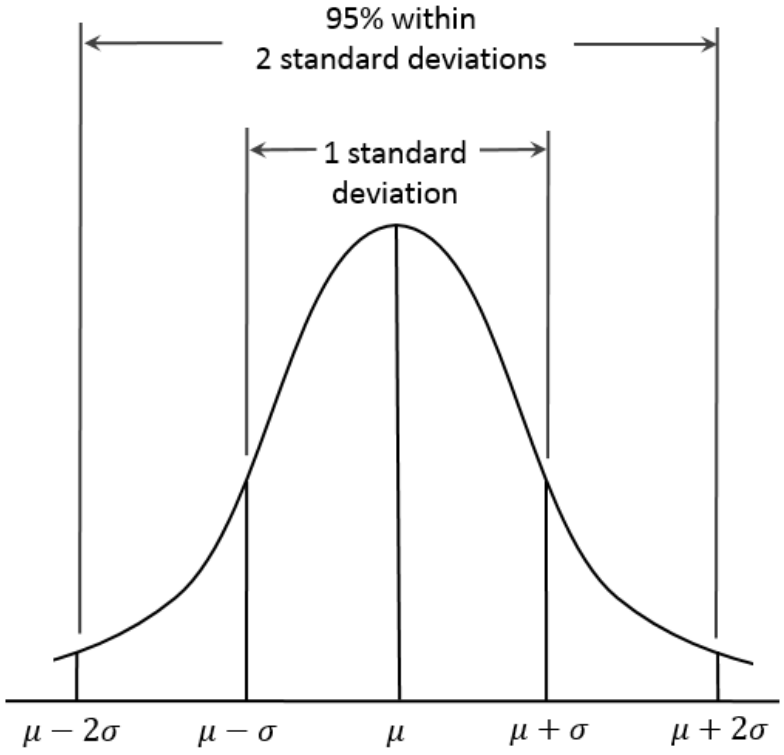
Normal distribution

One standard deviation *(one sigma)*



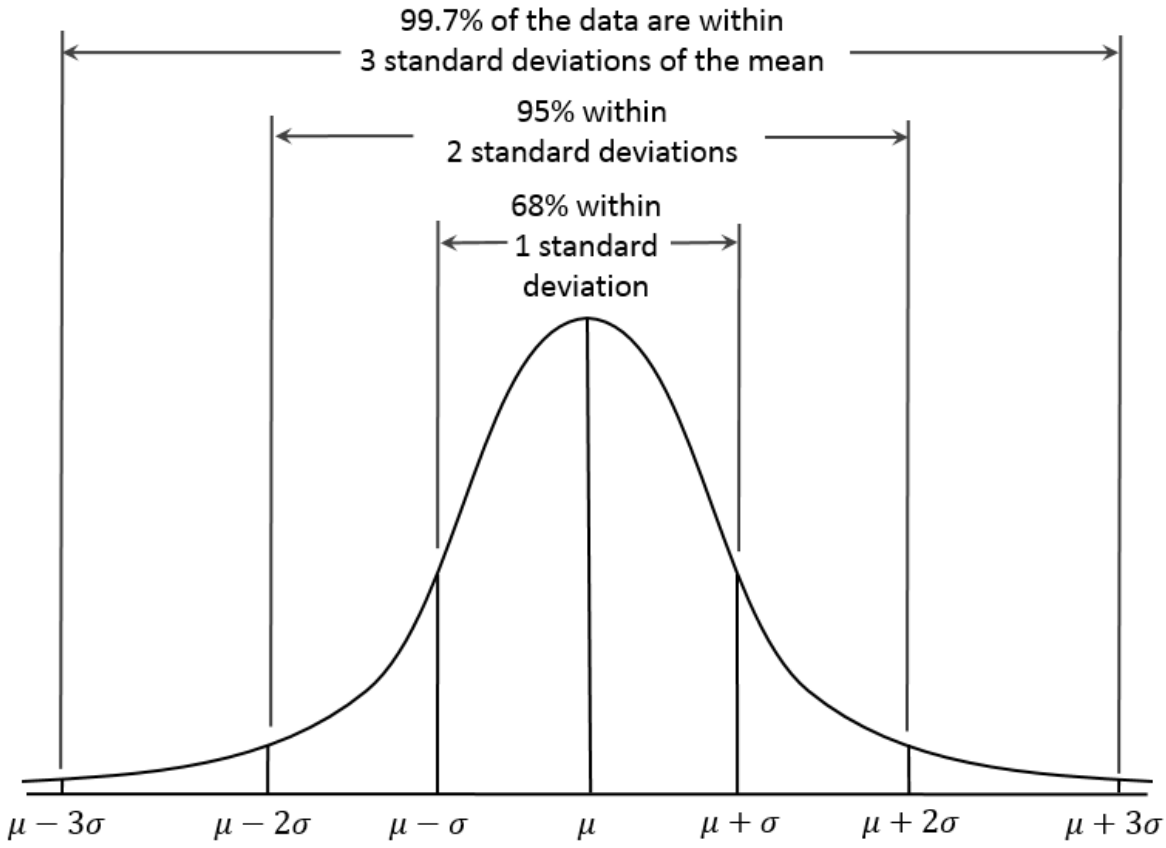
Normal distribution

Two standard deviations *(two sigma)*



Normal distribution

Three standard deviations *(three sigma)*



Is my distribution normal?

Visual aids

- Density plot
- Q-Q plot (quantile-quantile plot)
  `qqnorm()` or `ggplot(data) + aes(sample = x) + stat_qq()`

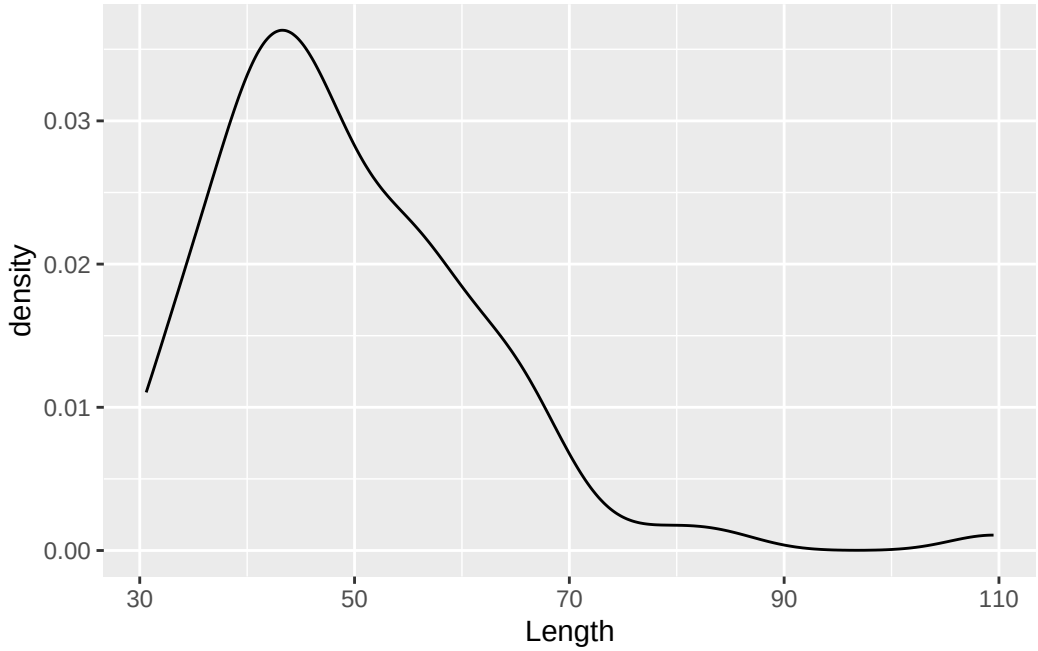
Statistical hypothesis test

- Shapiro-Wilk test
  `shapiro.text()`

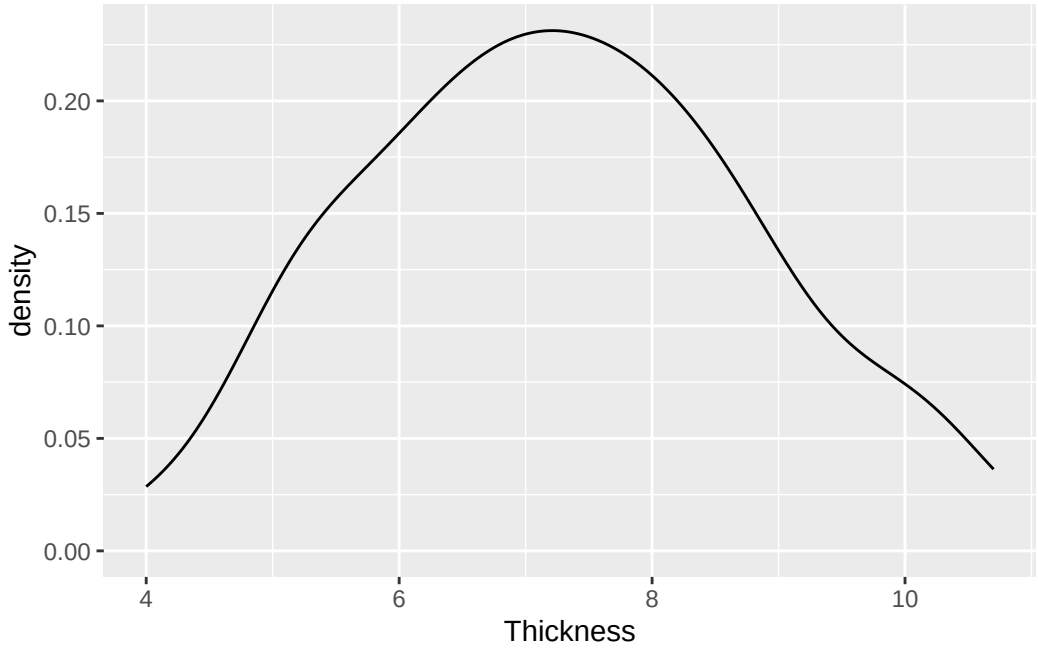
- Kolmogorov-Smirnov normality test

Q-Q plot

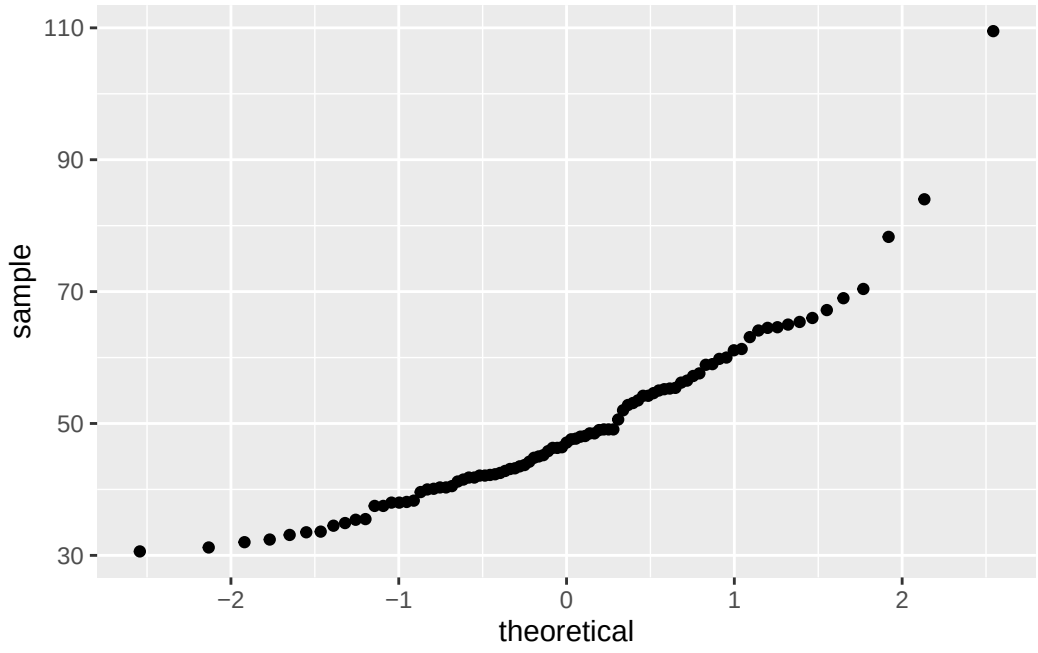
```
ggplot(dartpoints) + aes(x = Length) + geom_density()
```



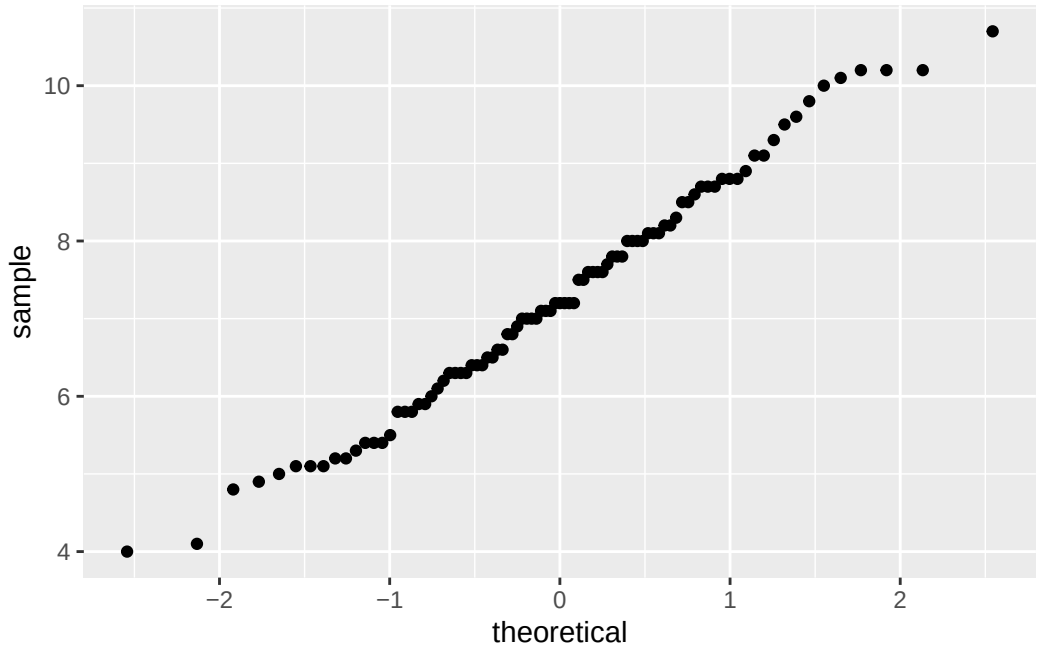

```
ggplot(dartpoints) + aes(x = Thickness) + geom_density()
```




```
ggplot(dartpoints) + aes(sample = Length) + stat_qq()
```



```
ggplot(dartpoints) + aes(sample = Thickness) + stat_qq()
```



Shapiro-Wilk normality test

- H_0 (null hypothesis): *Values fit normal distribution.*
- H_A (alternative hypothesis): *Values do not fit normal distribution.*
- **p-value**: *probability* of the event that observed values fit normal distribution
- p $>$ 0.05: Fail to reject null hypothesis.
- *Significance level* = 0.05 – Event occurs in less than 5% of cases

```
shapiro.test(dartpoints$Length)
```

```
    Shapiro-Wilk normality test

data:  dartpoints$Length
W = 0.90277, p-value = 4.852e-06
```

```r
shapiro.test(dartpoints$Thickness)
```

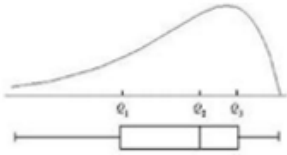
```
    Shapiro-Wilk normality test

data:  dartpoints$Thickness
W = 0.98623, p-value = 0.4559
```

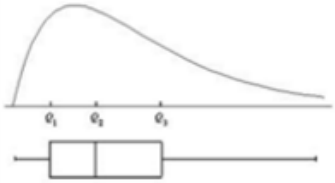
Other shapes of distributions

Normal distribution
(Hill/mound shapes, symmetric, Bell shaped curve)

Left skewed
(Tail is on the left hand side)



Right Skewed
(Tail is on the right hand side)



Multimodal
(There is more than one peak)