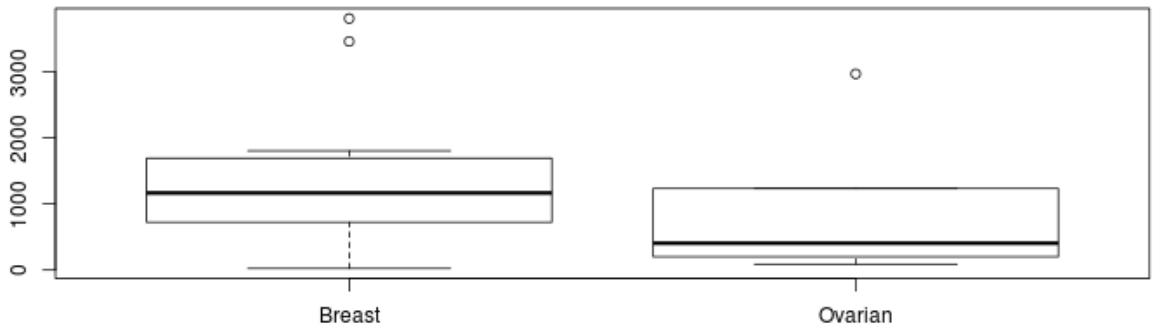
Choosing the correct test

Mark Dunning

4 November 2015

Dataset 1 `data1.csv`

Survival times of patients with Ovarian or Breast Cancer. Is there a difference in survival time between the two diseases?



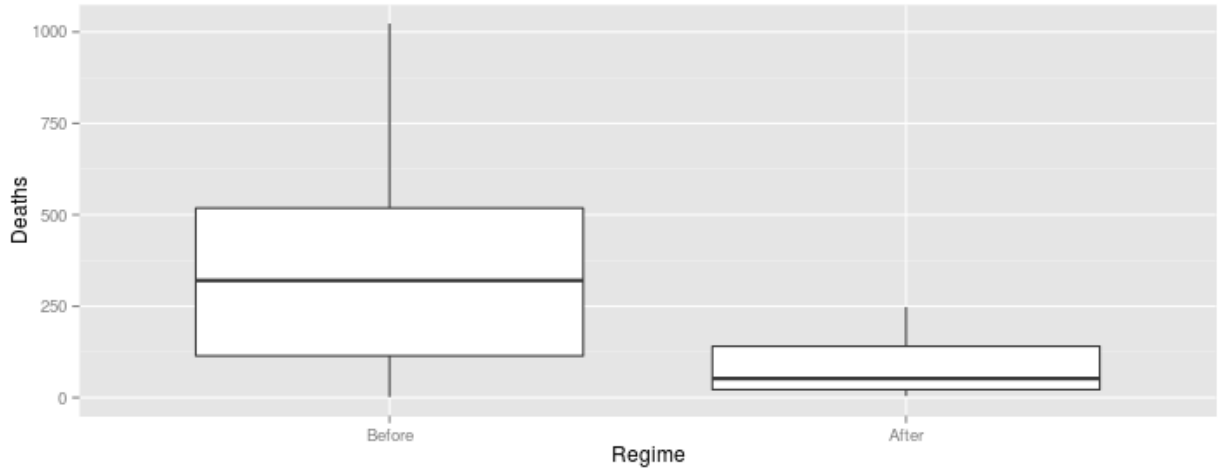

```
##
##  Wilcoxon rank sum test
##
## data:  Time by Disease
## W = 43, p-value = 0.3502
## alternative hypothesis: true location shift is not equal to 0
```

Dataset 2 `data2.csv`

In the history of data visualization, Florence Nightingale is best remembered for her role as a social activist and her view that statistical data, presented in charts and diagrams, could be used as powerful arguments for medical reform.

After witnessing deplorable sanitary conditions in the Crimea, she wrote several influential texts (Nightingale, 1858, 1859), including polar-area graphs (sometimes called "Coxcombs" or rose diagrams), showing the number of deaths in the Crimean from battle compared to disease or preventable causes that could be reduced by better battlefield nursing care.

Her Diagram of the Causes of Mortality in the Army in the East showed that most of the British soldiers who died during the Crimean War died of sickness rather than of wounds or other causes. It also showed that the death rate was higher in the first year of the war, before a Sanitary Commissioners arrived in March 1855 to improve hygiene in the camps and hospitals.




```
##
##  Wilcoxon rank sum test
##
## data:  Deaths by Regime
## W = 106, p-value = 0.02593
## alternative hypothesis: true location shift is greater than 0
```

Dataset 3: Gene expression `data3.csv`

The expression level of a gene was measured in a breast cancer cohort of ER negative and positive patients. Is the gene differentially-expressed between ER positive and negative patients?

```
##
##  Welch Two Sample t-test
##
## data:  Expression by ERStatus
## t = -38.746, df = 205.88, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.246953 -1.126198
## sample estimates:
## mean in group Negative mean in group Positive
##            -1.17388506             0.01269076
```

Dataset4: Sleep Data `data4.csv`

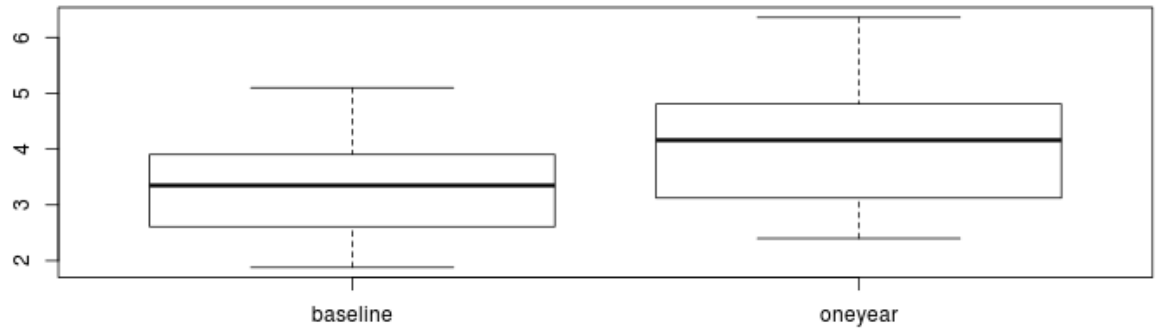
Data which show the effect of two soporific drugs (increase in hours of sleep compared to control) on 10 patients.

```
##
##  Paired t-test
##
## data:  sleep[, 1] and sleep[, 2]
## t = -4.0621, df = 9, p-value = 0.002833
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -2.4598858 -0.7001142
## sample estimates:
## mean of the differences
##                   -1.58
```

Dataset5: CD4 `data5.csv`

CD4 cells are carried in the blood as part of the human immune system. One of the effects of the HIV virus is that these cells die. The count of CD4 cells is used in determining the onset of full-blown AIDS in a patient. In this study of the effectiveness of a new anti-viral drug on HIV, 20 HIV-positive patients had their CD4 counts recorded and then were put on a course of treatment with this drug. After using the drug for one year, their CD4 counts were again recorded. The aim of the experiment was to show that patients taking the drug had increased CD4 counts which is not generally seen in HIV-positive patients.




```
##
##  Paired t-test
##
## data:  data[, 1] and data[, 2]
## t = -4.4908, df = 19, p-value = 0.0002504
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.1801882 -0.4298118
## sample estimates:
## mean of the differences
##                  -0.805
```

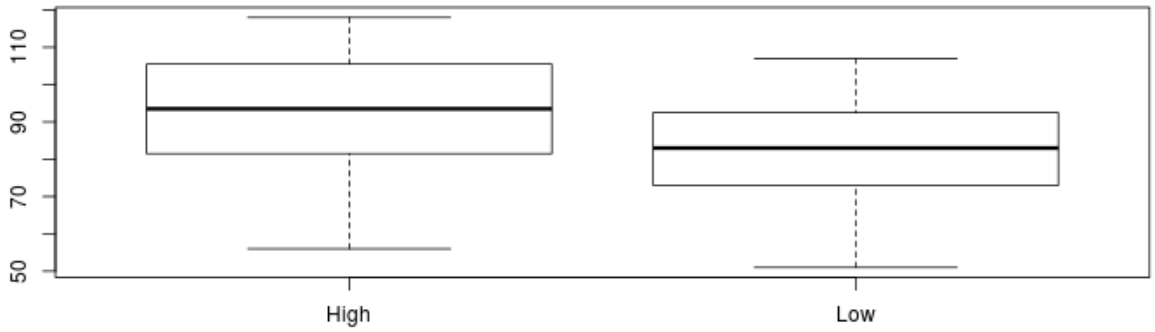
Dataset6: Birth Weight `data6.csv`

Risk Factors Associated with Low Infant Birth Weight

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  bwt by smoke
## W = 5249.5, p-value = 0.006768
## alternative hypothesis: true location shift is not equal to 0
```

Dataset7: Weight gain in Rats `data7.csv`

The data arise from an experiment to study the gain in weight of rats fed on four different diets, distinguished by amount of protein (low and high) and by source of protein (beef and cereal).



Dataset8: Colon cancer