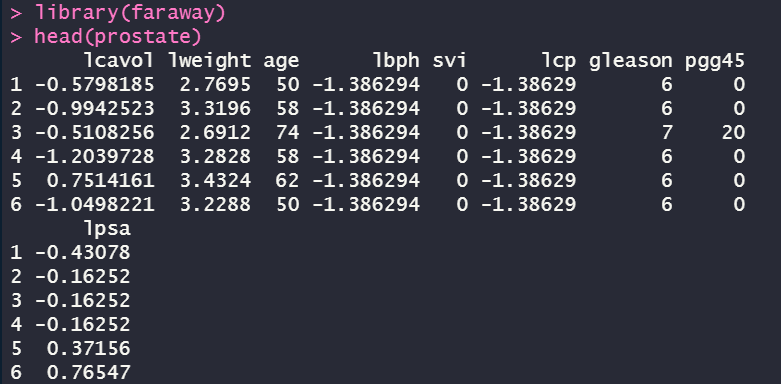
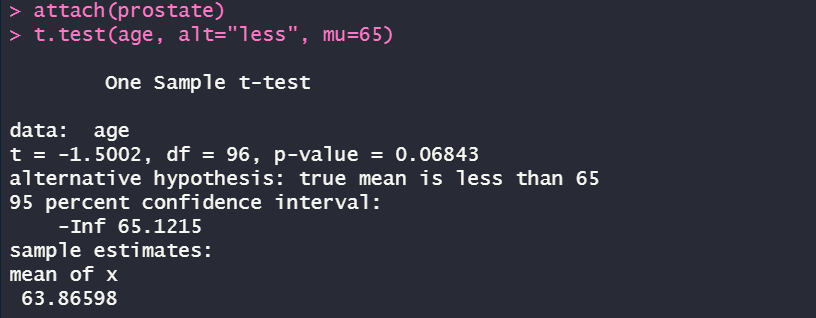
**STAT 40001/STAT 59800 Statistical Computing Fall 2020**

**Lab-10**

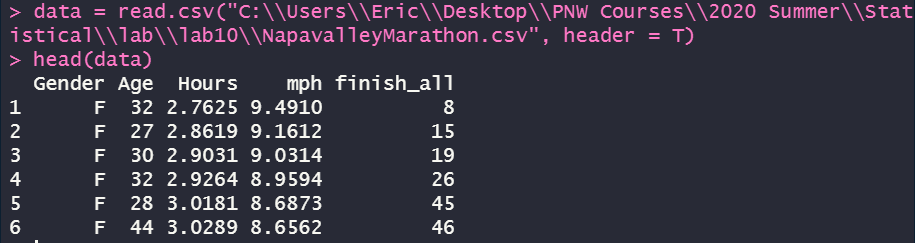
1. ***faraway*** package in R contains a data set ***prostate*** which describes 97 men with prostate cancer who were due to receive a radical prostatectomy. Test whether the participants are younger than 65 years.





*(From the above function, we can see that although the mean number of age is 63.86598 which is younger than 65 years old, the p-value is 0.06843, greater than the default significance level 0.05, so we can’t conclude that the participants are younger than 65 only based on mean value)*

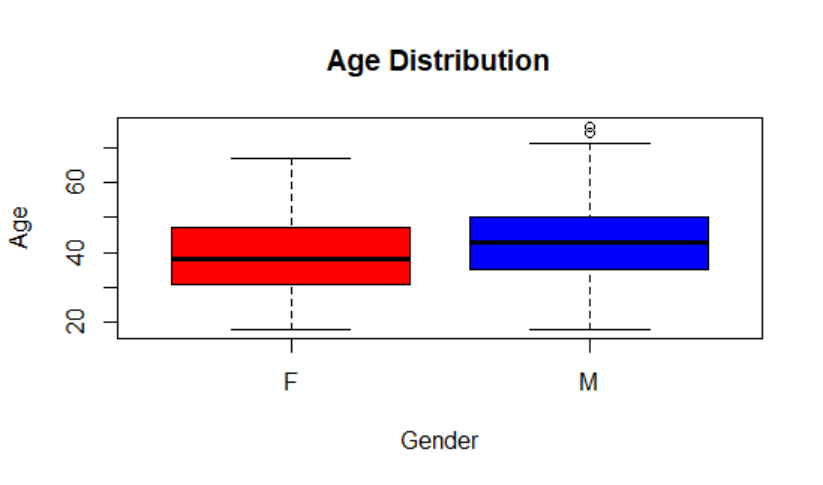
1. Napa Valley Marathon Times by Age and Gender for 2015 are provided with this assignment.
2. Import the data in R



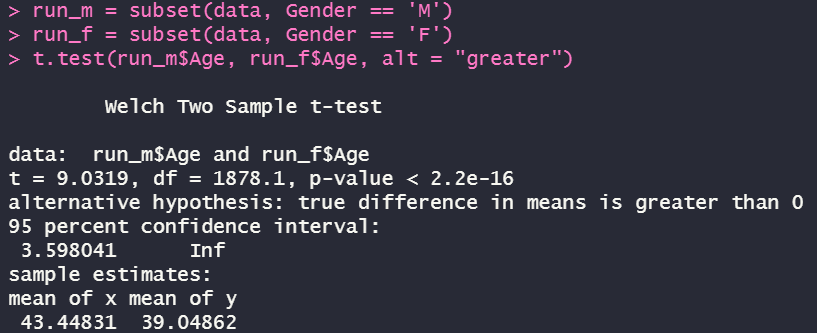
1. How many runners are older than 50 years of age?



1. Display the age distributions of the runner by gender.

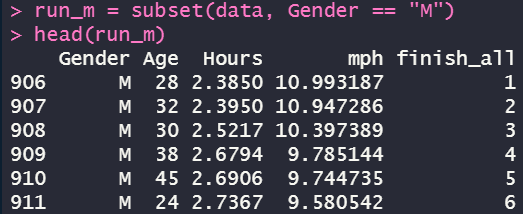
  


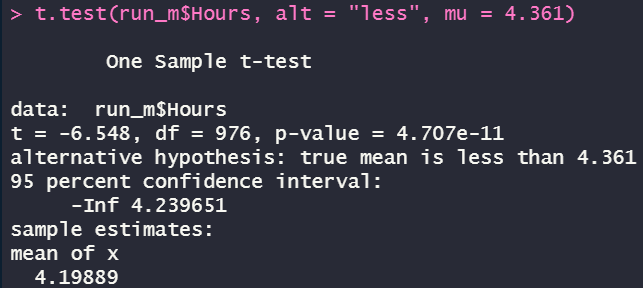
1. Are men older than women?



*(from the p-value is way less than 0.05, the default significance level, we can directly reject the Null Hypothesis because of no enough evidence and in favor of the alternative hypothesis that men are older than women)*

1. The average completion time for all runners is 4.361 hours. Test whether the completion time for men is lower than 4.361 hours.

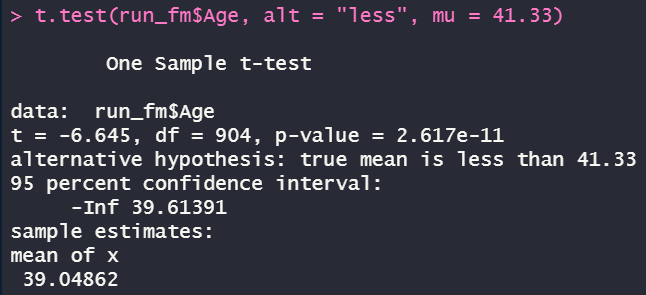




*(The p-value is pretty tiny that way less than 0.05 which in favor of our alternative hypothesis that the completion time for men is less than 4.361 hours)*

1. The average age for all runners is 41.33 years. Test whether women are younger than 41.33 years.



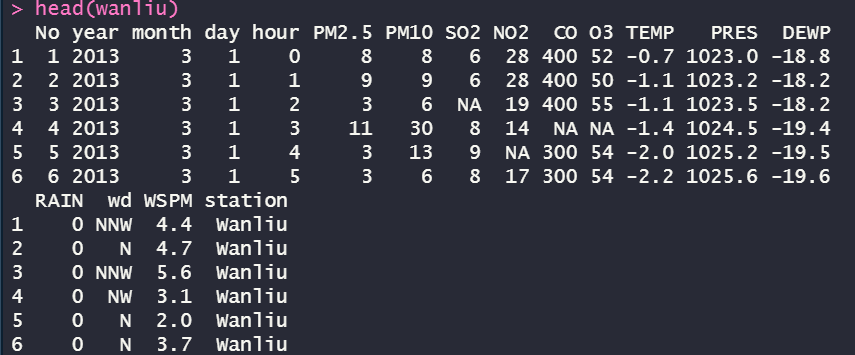


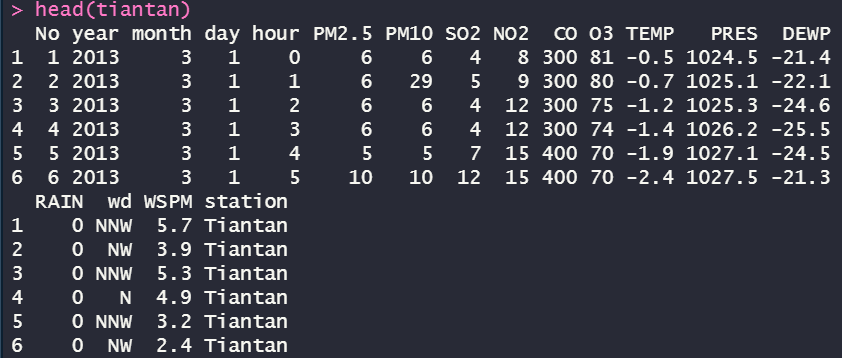
*(The p-value is extremely small which in favor of our alternative hypothesis that women are younger than 41.33 years )*

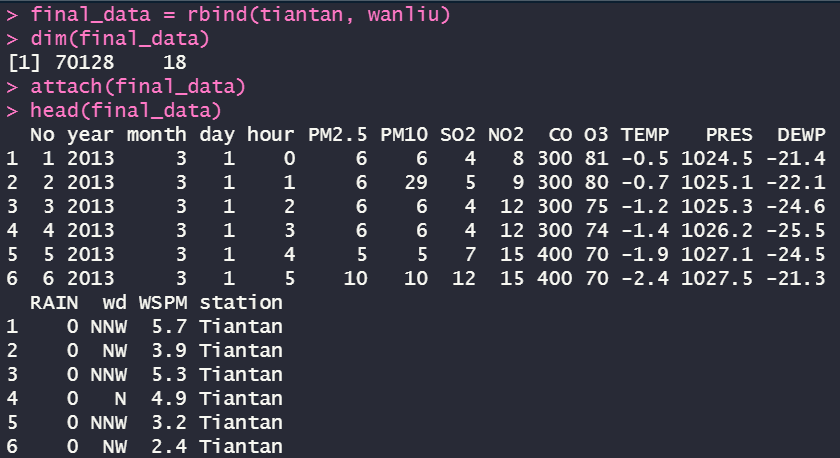
1. The data set **Beijing Multi-Site Air-Quality Data** provided in the UCI Machine Learning Repository includes hourly air pollutants data from 12 nationally-controlled air-quality monitoring sites. The air-quality data are from the Beijing Municipal Environmental Monitoring Center. The meteorological data in each air-quality site are matched with the nearest weather station from the China Meteorological Administration. The time period is from March 1st, 2013 to February 28th, 2017. Missing data are denoted as NA.

*Source: Zhang, S., Guo, B., Dong, A., He, J., Xu, Z. and Chen, S.X. (2017) Cautionary Tales on Air-Quality Improvement in Beijing. Proceedings of the Royal Society A, Volume 473, No. 2205, Pages 04-57.*

1. Import the data from Wanliu and Tiantan stations in R and create a single data frame.

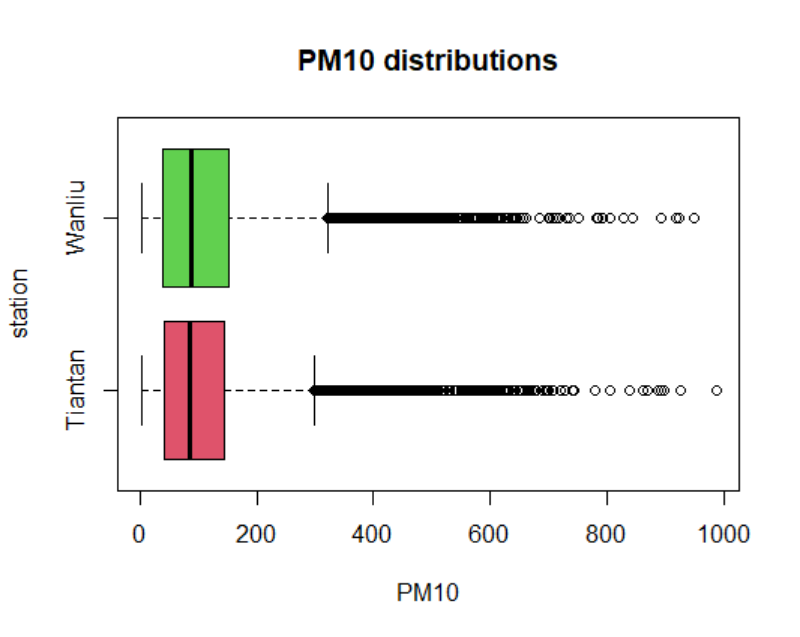




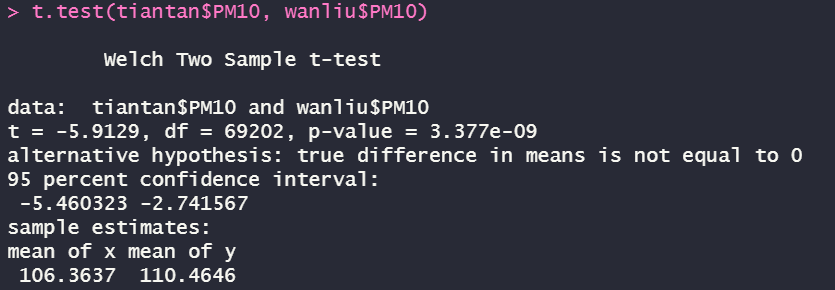


1. Display the PM10 values by creating a side-by-side box plot for both stations.





1. Test for significance difference in PM10 values in Wanliu and Tiantan.



*(The p-value is pretty tiny that we can directly reject Null Hypothesis and approve that there’s a significance difference between Wanliu and Tiantan)*