

Improving Statistical Competencies for Healthcare Undergraduate Students Using RStudio

Kylie Van Dyke and Frank Weber (Spring 2023)



UNIVERSITY OF MINNESOTA
ROCHESTER

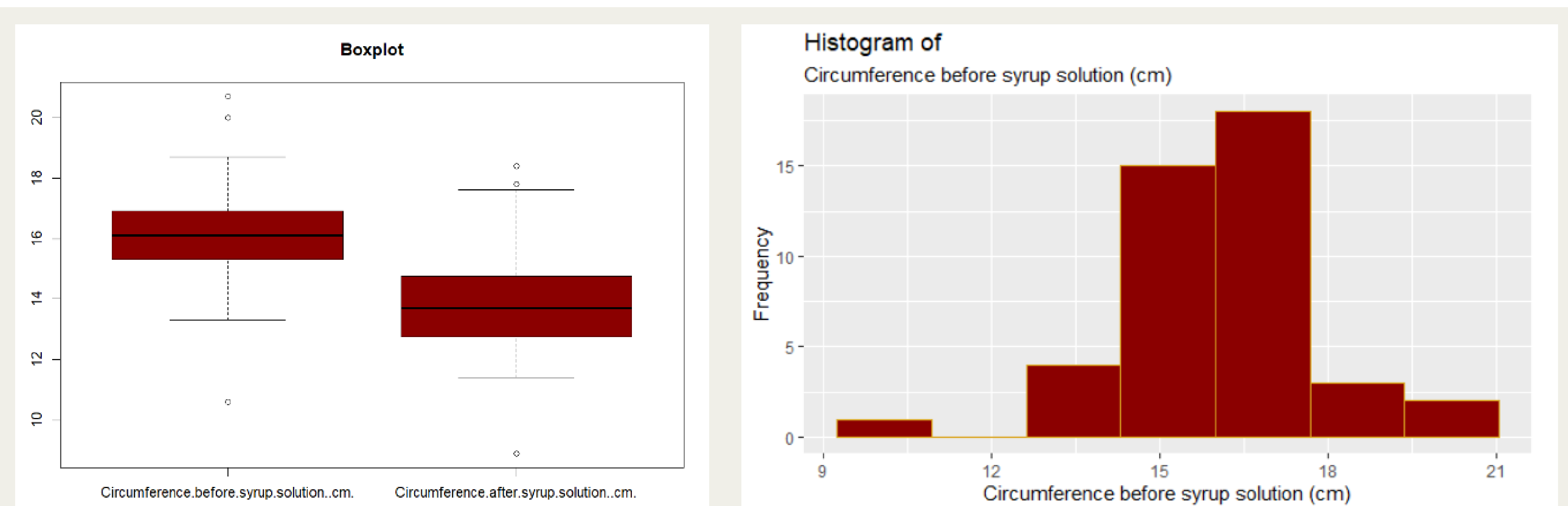
Abstract

Many undergraduate biological science courses offered at the University of Minnesota Rochester (UMR) utilize Excel for necessary data analysis. The courses always analyze at a parametric level even when the data are non-normally distributed. Each experiment typically employs a ubiquitous test, a derivative of a T-test. This often violates the foundational rules of statistics, since the students do not analyze or even visualize their data. They get the data, test it, and come out with a reported p-value. To properly shape an evidence based medicine mindset for their future, students must be encouraged to analyze the data and from there decide what level of test, if any, is appropriate. RStudio is an easily accessible integrated development environment (IDE) that incorporates the coding language R across all platforms. This presentation will describe the development of an R package. The package "Data Analysis for You-UMR" (DAY-UMR) will include tools to visualize data, test for normality, create figures, and get the most accurate p-value. DAY-UMR will make analyzing data straightforward for undergraduates and allow them easy access to multiple levels of testing. In this way it will provide more accurate and realistic experiences regarding research and data analysis. Overall, DAY-UMR will require more interaction with the data that will foster the desired evidence based medicine thinking in undergraduate students.

Introduction

R is a programming language that was made to analyze data and the platform RStudio was made to fully utilize the characteristics of R. Both have steadily been rising in popularity and serve as great way to analyze statistics [1]. Tools offered by the tidyverse, and other free to use packages, allow for greater interaction, transformation, and visualization of data compared to the statistical tools, such as Excel, most often used at UMR. The current tools used do not allow easy access to multiple levels of testing, prepare one for future technology in healthcare research, or promote an evidence based learning mindset. Therefore we have developed DAY-UMR.

Function 1: Outlier_Detect & Function 2: Check_Normal

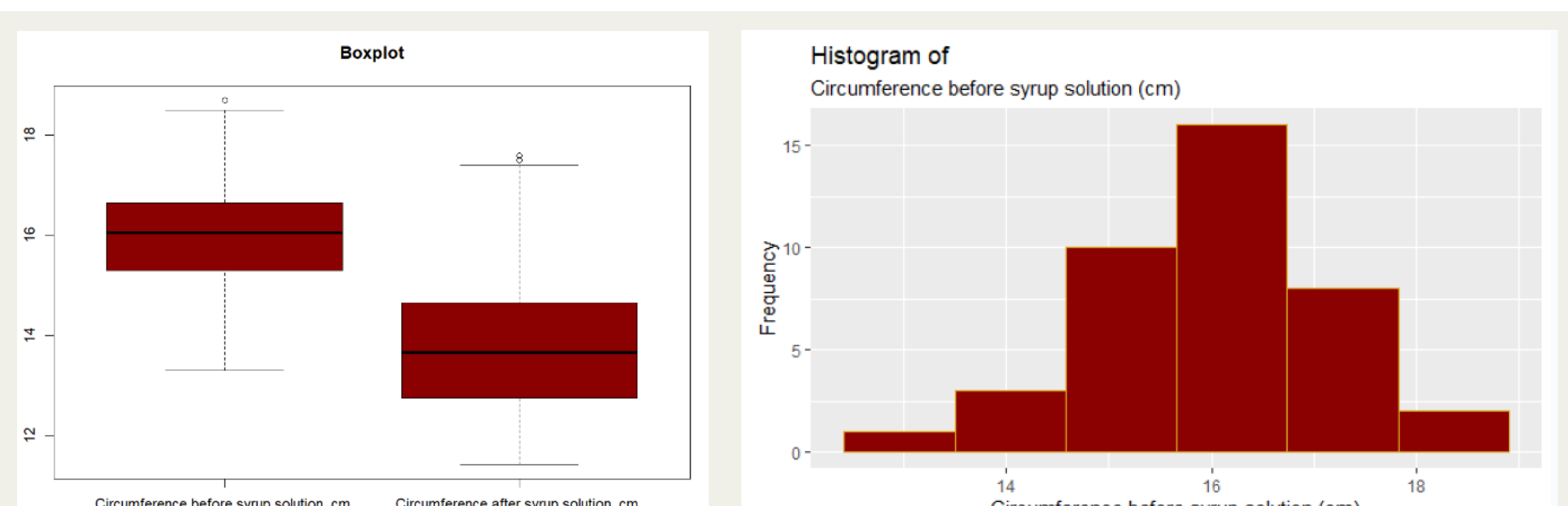


REMOVE==FALSE

```
The shapiro test for normality was conducted in addition to plotting a histogram.
The p-value is 0.0251 for Circumference before syrup solution (cm)
---
Significance(α) is set at 0.05 for this test.
Any value less than 0.05 indicates that the sample is not normally distributed
```

```
Outlier_Detect(data, col_c, col_x, paired,
REMOVE, plot_type)
```

```
Check_Normal(data, plot_type)
```

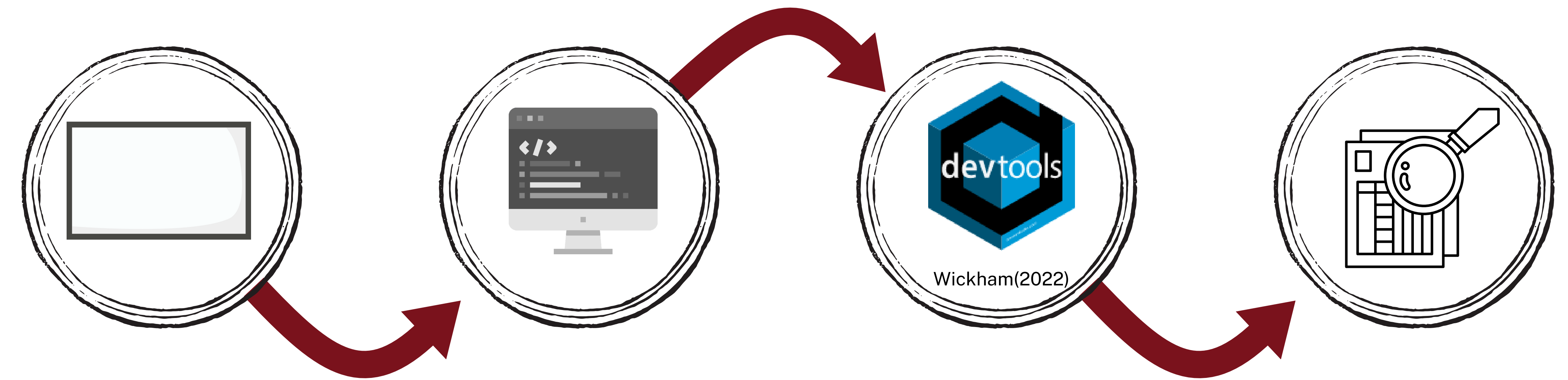


REMOVE==TRUE

```
The shapiro test for normality was conducted in addition to plotting a histogram.
The p-value is 0.7771 for Circumference before syrup solution (cm)
---
Significance(α) is set at 0.05 for this test.
Any value less than 0.05 indicates that the sample is not normally distributed
```

Conclusion

Excel is an easily accessible tool that most undergraduate students have some experience with but is difficult to use for students that don't have any experience with statistical analysis. Excel as it is currently implemented at the UMR, does not require students to understand their data. DAY-UMR will provide undergraduate students with a more accurate and realistic way to analyze data and result in more accurate p-values by allowing access to multiple levels of testing not offered in Excel. This higher level of interaction with data will foster critical thinking skills that will be required for a career in healthcare. Overall, DAY-UMR will improve how students think about and utilize data.



Purpose

Allow students to sharpen critical skills in statistical analysis such as visualizing data, removing outliers, executing appropriate levels of testing, and sharing results.

Development

- Pseudo Code
 - Walk through
 - What arguments?
- Coding
 - Learning Function implementation
 - Error after Error
- DevTools
 - New Tools
 - Learning
 - Future plans
- Testing
 - Different Datasets

Function 3: Test_Data

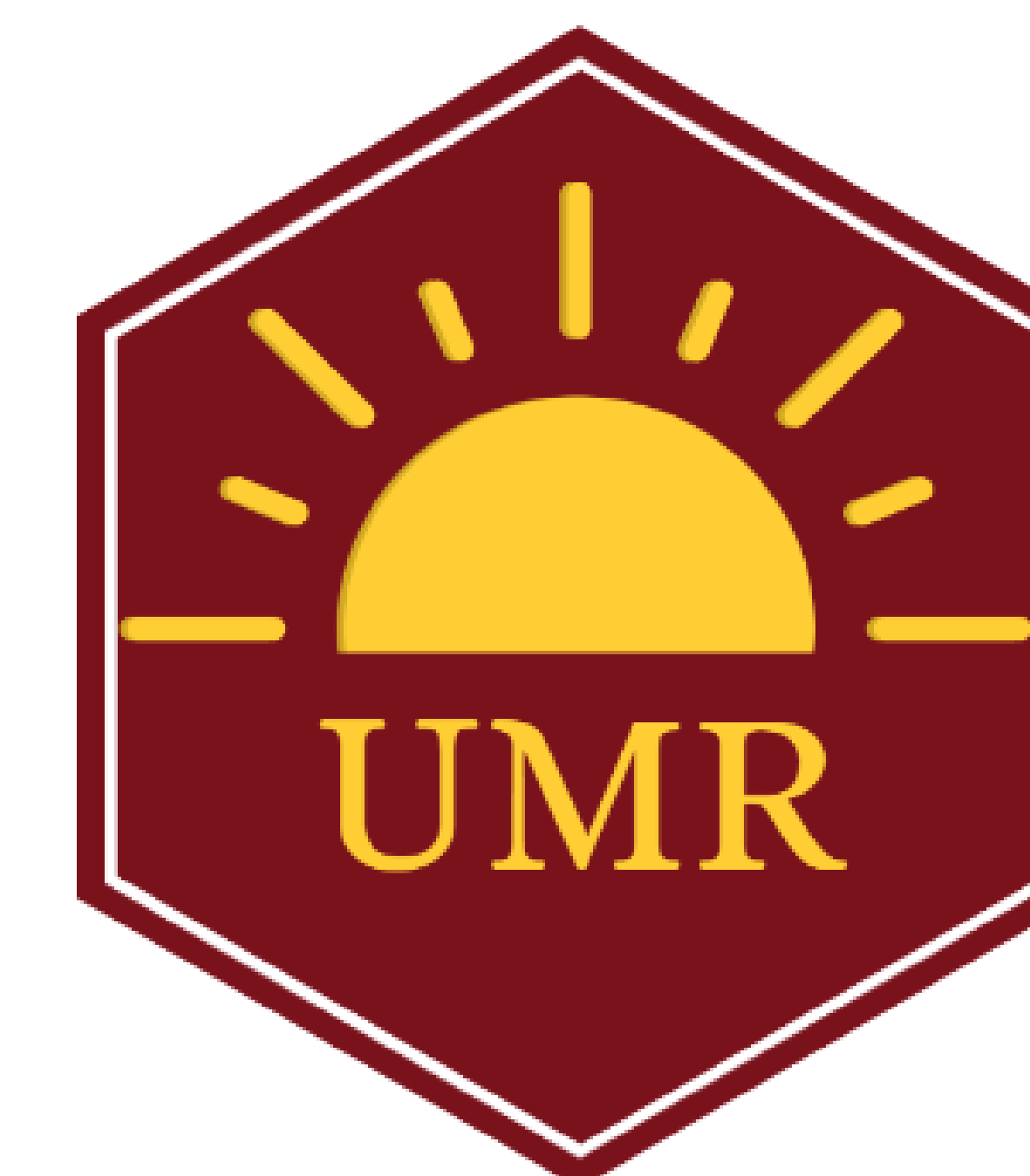
```
Test_Data(data, test_level, alternative)
```

Runs statistical test based on level (either parametric or nonparametric) given by the user and reports the p-value.

Function 4: Create_Figure

```
Create_Figure(data, test_level,
error_bars, figure_legend)
```

Produces bar graphs with error bars of means or medians depending on the level of the test either standard deviation or standard error given by the user with the ability to write a figure legend.



Potential logo for DAY-UMR



RStudio is a trademark of RStudio, PBC

Acknowledgements & References

Faculty Advisor: Dr. Abraham Ayebo

Hackenberger B. K. (2020). R software: unfriendly but probably the best. *Croatian medical journal*, 61(1), 66–68. <https://doi.org/10.3325/cmj.2020.61.66>

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686 <<https://doi.org/10.21105/joss.01686>>.

Wickham H, Hester J, Chang W, Bryan J (2022). *_devtools: Tools to Make Developing R Packages Easier_*. R package version 2.4.5, <<https://CRAN.R-project.org/package=devtools>>.