## horizontal line



Reflection

# 

# 

# 

# Activities

When I started learning R software, I felt a bit nervous. But after practicing, I found it to be a powerful tool for analyzing data. I want to share my experiences with four key areas: running R, importing/exporting files, doing calculations and managing variables, and creating graphs.

At first, I struggled to install R and RStudio on my laptop. I watched YouTube tutorials and asked friends for help. Once it was installed, I spent time getting familiar with the interface. I learned how to open scripts, run code line by line, and install packages like epitools and Hmisc. Running my first script successfully gave me confidence. I realized that patience is key when working with new software.

I imported the Health Survey for England 2011 dataset using the read\_sav() function. At first, I faced errors because the file path was incorrect. After fixing the path, the data loaded perfectly. I also learned to handle different file formats, though I mostly worked with .sav files. Exporting results was new to me, but I discovered I could save plots as images and tables as CSV files. This skill helped me share my findings easily.

Working with data taught me how to do calculations in R. For example, I calculated the percentage of drinkers using prop.table(), which showed 78.65% of people drink alcohol. I converted variables like "dnnow" to binary (1 = drinks, 0 = doesn’t drink) using ifelse(). Handling missing data was tricky, but using na.rm = TRUE in functions like summary() helped me avoid errors. I also learned to subset data, like removing rows with missing values for gender.

Creating visualizations was my favorite part. I made barplots to show drinking percentages by region using barplot(), and boxplots to compare heights and weights between genders. Initially, my plots lacked labels and titles, making them hard to understand. But after adding axes labels and colors, they became clearer. For instance, my boxplot showed males have a higher median height (173.3 cm) than females (160.4 cm). Seeing my data come to life in graphs was rewarding.

At first, I found it difficult to work with large datasets. I realized I needed to focus on specific parts of the data to answer my questions. I learned how to create sub-samples using R commands like subset(). For example, I created a sub-sample of drinkers by selecting rows where dnnow > 0. This helped me focus my analysis on people who consume alcohol. I also learned to filter data based on gender and other variables. Initially, I made mistakes with the syntax, but with practice, I became more confident. Creating sub-samples allowed me to explore specific groups within my data and made my analysis more targeted.

Understanding how to test hypotheses was a big step for me. I learned about the null hypothesis which means assuming no effect or difference and the alternative hypothesis which is assuming there is an effect or difference. For eg, I wanted to test if men and women differ in alcohol consumption. I set up a chi-square test using chisq.test() on a contingency table. The test gave me a p-value, which helped me decide whether to reject the null hypothesis. I found this process logical but needed to review the theory behind hypothesis testing to fully understand what I was doing. I also learned to calculate odds ratios, which gave me additional insights into the relationships in my data.

Interpreting p values was initially confusing. I learned that a p-value less than 0.05 typically indicates statistical significance. In my analysis, I found a p-value of less than 2.2e-16 for the gender and drinking relationship, which was strong evidence to reject the null hypothesis. This meant there was a significant difference in drinking habits between men and women. I also used p-values from other tests, like the Wilcoxon test for height and weight comparisons. Making decisions based on these values helped me draw meaningful conclusions from my data. I realized that statistical significance doesn't always mean practical importance, so I tried to interpret my results in context.

# 

# My Journey as an Independent Learner in Data Analysis

During this module, I've made great progress as an independent learner. I've learned to work on my own, solve problems, and complete tasks without much help from others. I want to share how I've grown in three key areas: performing data analysis, interpreting statistical results, and making decisions based on p-values. I'll also talk about the weekly activities I've done and how I've shared my knowledge with others.

## Performing Data Analysis and Interpreting Results

At the beginning of the module, I had some knowledge of R, but I wasn't very confident. I decided to dive deep into learning R by practicing regularly and working on real datasets. I used the Health Survey for England 2011 dataset to analyze alcohol consumption patterns. I started by importing the data using the read\_sav() function and then spent time exploring it. I learned how to calculate basic statistics like mean, median, and mode for variables such as household size, BMI, and age. For example, I found that the mean household size was 2.85, and the median age was 42 years.

I also learned how to create sub-samples using the subset() function. I created a sub-sample of drinkers by selecting rows where dnnow > 0. This helped me focus my analysis on specific groups within the data. I faced some challenges with syntax and data structure, but I was able to overcome them by referring to online resources and documentation.

## Making Decisions Based on P-values

Understanding hypothesis testing was a big milestone for me. I learned about the null hypothesis (assuming no effect or difference) and the alternative hypothesis (assuming there is an effect or difference). I performed a chi-square test to determine if there's a significant association between gender and alcohol consumption using the chisq.test() function. The test resulted in a p-value of less than 2.2e-16, which strong evidence to reject the null hypothesis. This meant that there's a significant difference in drinking habits between men and women.

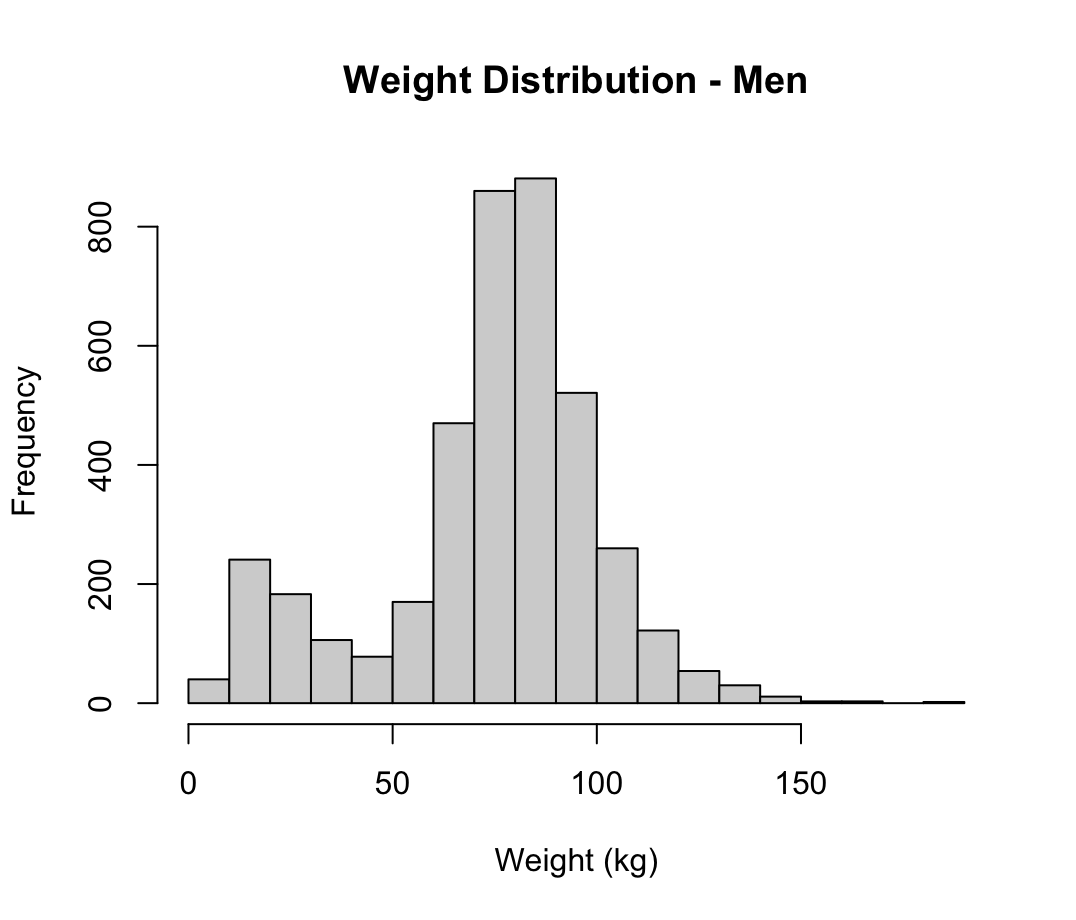
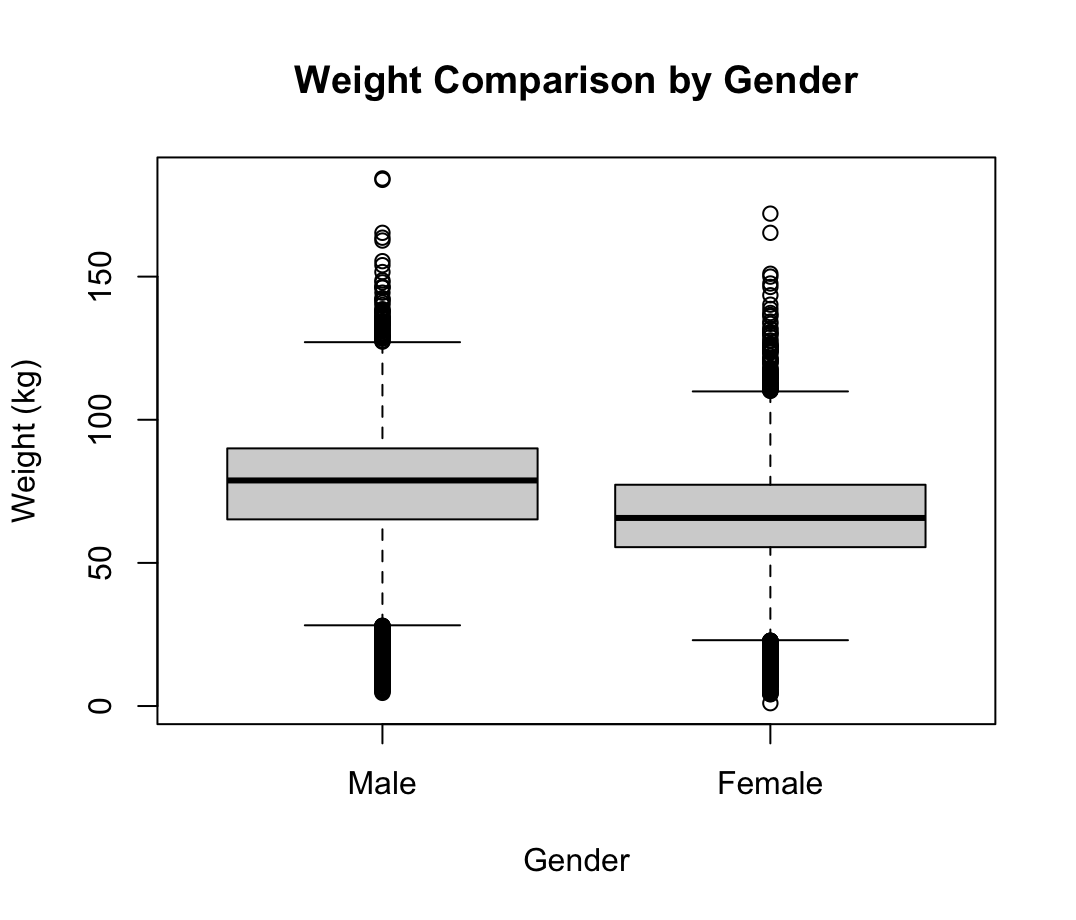
I also used the Wilcoxon test to compare height and weight between genders. Both tests showed significant differences, with p-values less than 2.2e-16. These results helped me draw meaningful conclusions about the data. I realized that interpreting p-values is crucial for making informed decisions based on statistical analysis.

## Weekly Activities and Knowledge Sharing

Throughout the module, I engaged in various weekly activities to reinforce my learning. I practiced coding in R regularly, worked on different datasets, and participated in online forums to discuss challenges and solutions with fellow learners. I tried to maintain an e-portfolio on github where I could document my progress, save my R scripts, and store visualizations I created - but I got stuck at one point.

Screenshots of my R code for importing data and performing basic statistics.

Visualizations like barplots and boxplots that I created to compare different variables.



Boxplot and barplots for comparing Weight by Gender and weights distribution

Results of hypothesis tests, including p-values and odds ratios.

### Research and Further Learning

To enhance my understanding of alcohol consumption patterns in the UK, I conducted some research. I read several academic papers and reports, including:

1. "Associations between socio-economic factors and alcohol consumption: A population survey of adults in England" by Beard et al. (2019)
2. "A rapid evidence review of the effectiveness and cost-effectiveness of alcohol control policies: an English perspective" by Burton et al. (2017)
3. "Statistics on Alcohol, England 2019" by NHS Digital

These resources provided good insights into the socio economic factors influencing alcohol consumption along with the effectiveness of various control policies. This research helped me to draw more informed conclusions.

# 

# My Emotional Journey Learning R

Working on this module has been an emotional journey for me. I felt nervous at times and unsure, especially when I first started learning R. Coding has always been overwhelming for me. The software was new to me, and I was worried I wouldn't be able to understand it. But as I worked through the tasks, I began to feel more confident. I realized that with practice and patience, I could learn new skills.

## Feelings and Emotions

At the beginning of the module, I felt a bit scared. R seemed complicated, and I wasn't sure if I could do it. I remember thinking, "This is going to be really hard." But I also felt eager to learn. I wanted to understand data analysis and use R to answer interesting questions.

When I faced challenges, like trying to understand hypothesis testing or fix errors in my code, I sometimes felt frustrated. For example, I spent a lot of time trying to get the read\_sav() function to work properly. At first, I kept getting error messages because the file path was incorrect. I felt like giving up, but then when I found that the problem was a very small one, I laughed at my frustration.

These emotions affected how I approached my work. When I felt confident, I was more willing to try new things and experiment with different functions in R. When I was frustrated, I sometimes avoided working on the tasks. But I learned to manage these emotions by taking breaks and coming back to the work with a fresh mind.

To overcome the challenges, I used several strategies. First, I broke down complex tasks into smaller steps. Instead of trying to do everything at once, I focused on one thing at a time. For example, when creating visualizations, I first learned to make a simple barplot and then added labels and colors.

I also took help when I needed it. I watched online tutorials, read documentation, and asked classmates for advice. For instance, when I was struggling with hypothesis testing, I found a helpful tutorial on YouTube that explained the concepts in a simple way. I also referred to academic literature, such as "An R Companion to Applied Regression" by Fox and Weisberg (2019), which provided valuable guidance on statistical methods and R programming.

## Group Learning and Discussions

Discussing my work with others had a big impact on my learning. When I shared my code and results with classmates, they gave me feedback and suggested improvements. For example, one classmate showed me a more efficient way to create sub-samples using the subset() function. These discussions helped me see different perspectives and approaches to solving problems.

At times, group discussions could be a bit discouraging. When others seemed to understand concepts faster than I did, I felt a bit left behind. But I reminded myself that everyone learns at their own pace, and it's okay to take time to understand things.

## Reviewing Others' Views and Literature

, I read "Associations between socio-economic factors and alcohol consumption: A population survey of adults in England" by Beard et al. (2019). This paper helped me understand how socio-economic factors influence alcohol consumption, which added depth to my analysis of the Health Survey for England 2011 dataset.

By reviewing literature and discussing my work with others, I was able to critically evaluate my own behavior and approaches. I realized that there are often multiple ways to solve a problem in R, and being open to different methods can improve my work.

# 

# Learning Theory: Growth Mindset

During this module, I've focused on applying the Growth Mindset theory to my learning process. This theory suggests that abilities can be developed through dedication and hard work. I've embraced challenges and viewed mistakes as opportunities to learn and grow. and learn from them.

## Application of Growth Mindset

Embracing Challenges: When I first started learning R, I felt nervous about tackling complex tasks like hypothesis testing and creating visualizations. But with the Growth Mindset, I viewed these challenges as chances to improve my skills. I spent extra time practicing and experimenting with different functions in R.

Viewing Mistakes as Learning Opportunities: There were times when I made errors in my code or misunderstood statistical concepts. Instead of getting discouraged, I used these mistakes as opportunities to learn. For example, when I initially struggled with understanding p-values, I reviewed resources and practiced until I grasped the concept.

Continuous Improvement: The Growth Mindset encouraged me to keep refining my skills. After completing an analysis, I would review my work, identify areas for improvement, and make changes. This iterative process helped me become more proficient in R and data analysis.

Persistence: There were moments when I wanted to give up, especially when faced with difficult concepts. But the Growth Mindset motivated me to persevere. I reminded myself that learning takes time and effort, and that I could improve with practice.

# References

Dweck, C. S. (2016). Carol Dweck Revisits the "Growth Mindset." Education Week, 35(8), 8-9.

Yeager, D. S., & Dweck, C. S. (2012). Mindsets That Promote Resilience: When Students Believe That Personal Characteristics Can Be Developed. Educational Psychologist, 47(4), 302-314.