STA 3100 Programming With Data in R: Data Analysis

#Data Analysis in R

##Introduction

Data Analysis is the process of examining data sets in order to draw conclusions about the information they contain. It involves using techniques from statistics, machine learning, and computer science to extract useful insights from data. In this lecture, we will learn how to use the R programming language to perform data analysis.

##Getting Started

Before we can start analyzing data, we need to make sure that we have the necessary tools. We will be using the R programming language, which is available for free from the Comprehensive R Archive Network (CRAN). We also need to install any additional packages that we may need for our analysis.

##Exploratory Data Analysis

Exploratory Data Analysis (EDA) is the process of exploring a dataset to gain insight into its structure and content. It involves using graphical and numerical techniques to summarize the data and identify patterns or relationships between variables.

###Coding Example 1: Summary Statistics

```
Start of Code
# Load the necessary packages
library(tidyverse)
# Read in the dataset
data <- read_csv("dataset.csv")
# Calculate summary statistics
summary_stats <- data %>%
    summarise_all(funs(mean, sd, min, max))
# Print the summary statistics
print(summary_stats)
End of Code
```

###Practice Multiple Choice Questions

Q: What is Exploratory Data Analysis?

A: Exploratory Data Analysis (EDA) is the process of exploring a dataset to gain insight into its structure and content. It involves using graphical and numerical techniques to summarize the data and

identify patterns or relationships between variables.