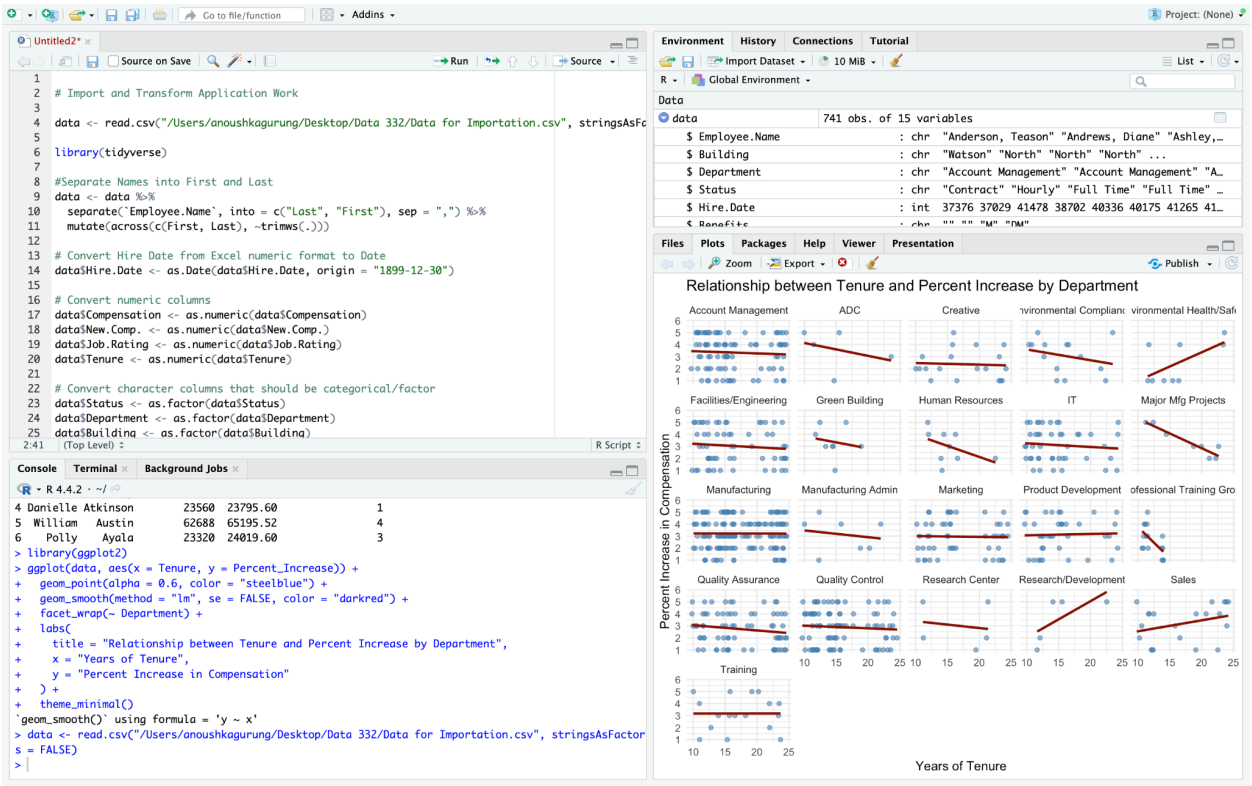
Import and Transform Application Work

```
# Import and Transform Application Work

data <- read.csv("/Users/anoushkagurung/Desktop/Data 332/Data for Importation.csv",
stringsAsFactors = FALSE)

library(tidyverse)

#Separate Names into First and Last
data <- data %>%
  separate(`Employee.Name`, into = c("Last", "First"), sep = ",") %>%
  mutate(across(c(First, Last), ~trimws(.)))

# Convert Hire Date from Excel numeric format to Date
data$Hire.Date <- as.Date(data$Hire.Date, origin = "1899-12-30")

# Convert numeric columns
data$Compensation <- as.numeric(data$Compensation)
data$New.Comp. <- as.numeric(data$New.Comp.)
data$Job.Rating <- as.numeric(data$Job.Rating)
data$Tenure <- as.numeric(data$Tenure)

# Convert character columns that should be categorical/factor
data$Status <- as.factor(data$Status)
data$Department <- as.factor(data$Department)
data$Building <- as.factor(data$Building)
data$Benefits <- as.factor(data$Benefits)

# Calculate Years of Tenure

data$Tenure <- round(as.numeric(difftime(Sys.Date(), data$Hire.Date, units = "days")) / 365, 2)


# Format Compensation and New Compensation into US Dollars
data <- data %>%
  mutate(
    Compensation_USD = paste0("$", formatC(Compensation, format = "f", big.mark = ",", digits
= 2)),
```

```
    New_Compensation_USD = paste0("$", formatC(New.Comp., format = "f", big.mark = ",",
digits = 2))
  )


# Calculate the percent increase between Compensation and New Compensation
data <- data %>%
  mutate(
    Percent_Increase = round(((New.Comp. - Compensation) / Compensation) * 100, 2)
  )

data %>%
  select(First, Last, Compensation, New.Comp., Percent_Increase) %>%
  head()

# Demonstrate for each department the relationship between Tenure and Compensation
Increase using scatter plots and regression

library(ggplot2)

ggplot(data, aes(x = Tenure, y = Percent_Increase)) +
  geom_point(alpha = 0.6, color = "steelblue") +
  geom_smooth(method = "lm", se = FALSE, color = "darkred") +
  facet_wrap(~ Department) +
  labs(
    title = "Relationship between Tenure and Percent Increase by Department",
    x = "Years of Tenure",
    y = "Percent Increase in Compensation"
  ) +
  theme_minimal()
```

Anoushka 1


Untitled2*
Source on Save
Run
Source
# Import and Transform Application Work
data <- read.csv("/Users/anoushkagurung/Desktop/Data 332/Data for Importation.csv", stringsAsFc
library(tidyverse)
#Separate Names into First and Last
data <- data %>%
separate(`Employee.Name`, into = c("Last", "First"), sep = ",") %>%
mutate(across(c(First, Last), ~trimws(.)))
# Convert Hire Date from Excel numeric format to Date
data$Hire.Date <- as.Date(data$Hire.Date, origin = "1899-12-30")
Convert numeric columns
data$Compensation <- as.numeric(data$Compensation)
data$New.Comp. <- as.numeric(data$New.Comp.)
data$Job.Rating <- as.numeric(data$Job.Rating)
data$Tenure <- as.numeric(data$Tenure)
Convert character columns that should be categorical/factor
data$Status <- as.factor(data$Status)
data$Department <- as.factor(data$Department)
data$Building <- as.factor(data$Building)
2:41
(Top Level)
R Script
Console
Terminal
Background Jobs
R 4.4.2 · ~/
4 Danielle Atkinson 23560 23795.60 1
5 William Austin 62688 65195.52 4
6 Polly Ayala 23320 24019.60 3
> library(ggplot2)
> ggplot(data, aes(x = Tenure, y = Percent_Increase)) +
+ geom_point(alpha = 0.6, color = "steelblue") +
+ geom_smooth(method = "lm", se = FALSE, color = "darkred") +
+ facet_wrap(~ Department) +
+ labs(
+ title = "Relationship between Tenure and Percent Increase by Department",
+ x = "Years of Tenure",
+ y = "Percent Increase in Compensation"
+) +
+ theme_minimal()
`geom_smooth()` using formula = 'y ~ x'
> data <- read.csv("/Users/anoushkagurung/Desktop/Data 332/Data for Importation.csv", stringsAsFactor
s = FALSE)
>
Environment
History
Connections
Tutorial
Import Dataset
10 MiB
List
R
Global Environment
Data
data
741 obs. of 15 variables
$ Employee.Name : chr "Anderson, Teason" "Andrews, Diane" "Ashley,…
$ Building : chr "Watson" "North" "North" "North" ...
$ Department : chr "Account Management" "Account Management" "A…
$ Status : chr "Contract" "Hourly" "Full Time" "Full Time" …
$ Hire.Date : int 37376 37029 41478 38702 40336 40175 41265 41…
Files
Plots
Packages
Help
Viewer
Presentation
Zoom
Export
Publish
Relationship between Tenure and Percent Increase by Department
Account Management
ADC
Creative
Facilities/Engineering
Green Building
Human Resources
IT
Major Mfg Projects
Manufacturing
Manufacturing Admin
Marketing
Product Development
Quality Assurance
Quality Control
Research Center
Research/Development
Sales
Training
Percent Increase in Compensation
Years of Tenure