

STAT630 HW 6

David Teng

Due Tuesday, November 5th by 11:59pm

```
chs <- read.table("data/chsData.txt", header = TRUE)
```

Part 0: Literature Review

(1.) Reputable Resources and Findings:

Mayo Clinic: An informative page on how weight gain is associated with elevated blood pressure, discussing the strain that extra weight places on the cardiovascular system and potential solutions for weight management to control SBP.

Link to article: <https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/symptoms-causes/syc-20373410>

National Institutes of Health (NIH) Study: This research paper delves into the correlation between abdominal obesity and higher blood pressure, explaining physiological mechanisms like increased cardiac output that lead to higher SBP.

Link to study: <https://pmc.ncbi.nlm.nih.gov/articles/PMC4140623/>

(2.) Relationship Between Weight and SBP: Based on the articles, I believe there is a positive correlation between weight and SBP. Individuals with higher body weight are more likely to have elevated SBP due to factors like increased cardiac output and vascular resistance. This indicates that maintaining a healthy weight is important for blood pressure management.

Part 1: Data Cleaning

(3.) (a)

```
# Convert categorical variables to factors
chs$clinic <- factor(chs$clinic, levels = 1:4,
                    labels = c("Sacramento", "Forsyth", "Washington", "Pittsburgh"))
chs$season <- factor(chs$season, levels = 1:4,
                    labels = c("summer", "fall", "winter", "spring"))
chs$arthritis <- factor(chs$arthritis, levels = c(0, 1), labels = c("none", "arthritis"))
chs$diabetes <- factor(chs$diabetes, levels = c(1, 2, 3),
                     labels = c("none", "borderline", "diabetes"))
chs$income <- factor(chs$income, levels = 1:8,
                    labels = c("<5k", "5k-8k", "8k-12k", "12k-16k",
                               "16k-24k", "24k-35k", "35k-50k", ">50k"))
chs$exint0 <- factor(chs$exint0, levels = 0:3,
                    labels = c("no exercise", "low intensity", "moderate intensity", "high intensity"))
```

(3.) (b)

```
library(dplyr)
# Create a new binary variable for high systolic blood pressure (sbp140)
```

```
chs <- chs %>%
  mutate(sbp140=ifelse(sbp>=140,1,0))
```

Part 2: Exploratory Data Analysis

(4.) (a) Count Missing Values

```
knitr::kable(apply(chs,2, function(x) sum(is.na(x))))
```

	x
clinic	0
initdate	0
season	0
gender	0
age	0
weight	7
weight50	88
grade	6
arth	33
sbp	7
pkyrs	67
diab	17
income	157
exint0	2
block0	25
kcal0	4
sbp140	7

(4.) (b)

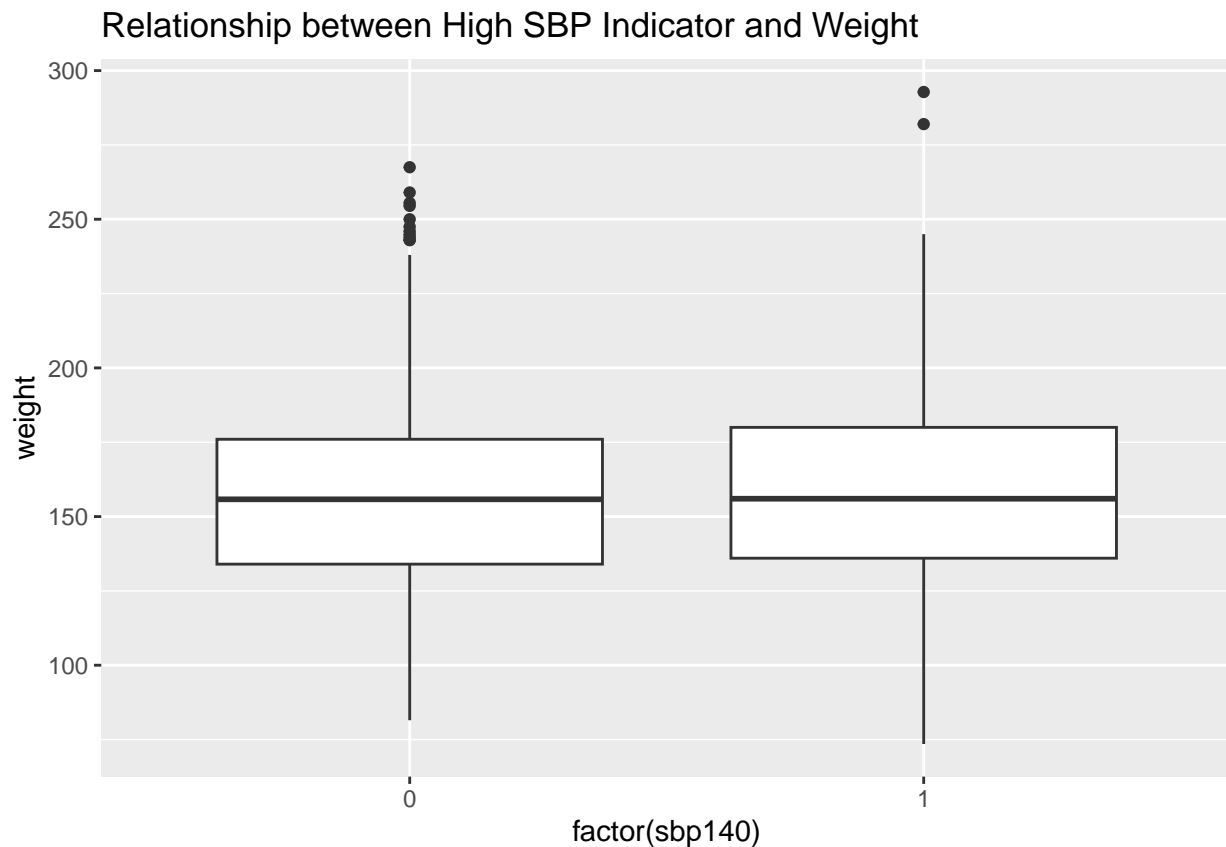
Removing missing values makes the dataset easier to handle and prevents problems during analysis. It keeps things uniform and avoids gaps that could throw off our results. While there might be some bias, having a clean, complete dataset can make our conclusions clearer and more reliable.

```
# Remove rows with missing values
chs_clean <- na.omit(chs)
```

(5.)

```
library(ggplot2)

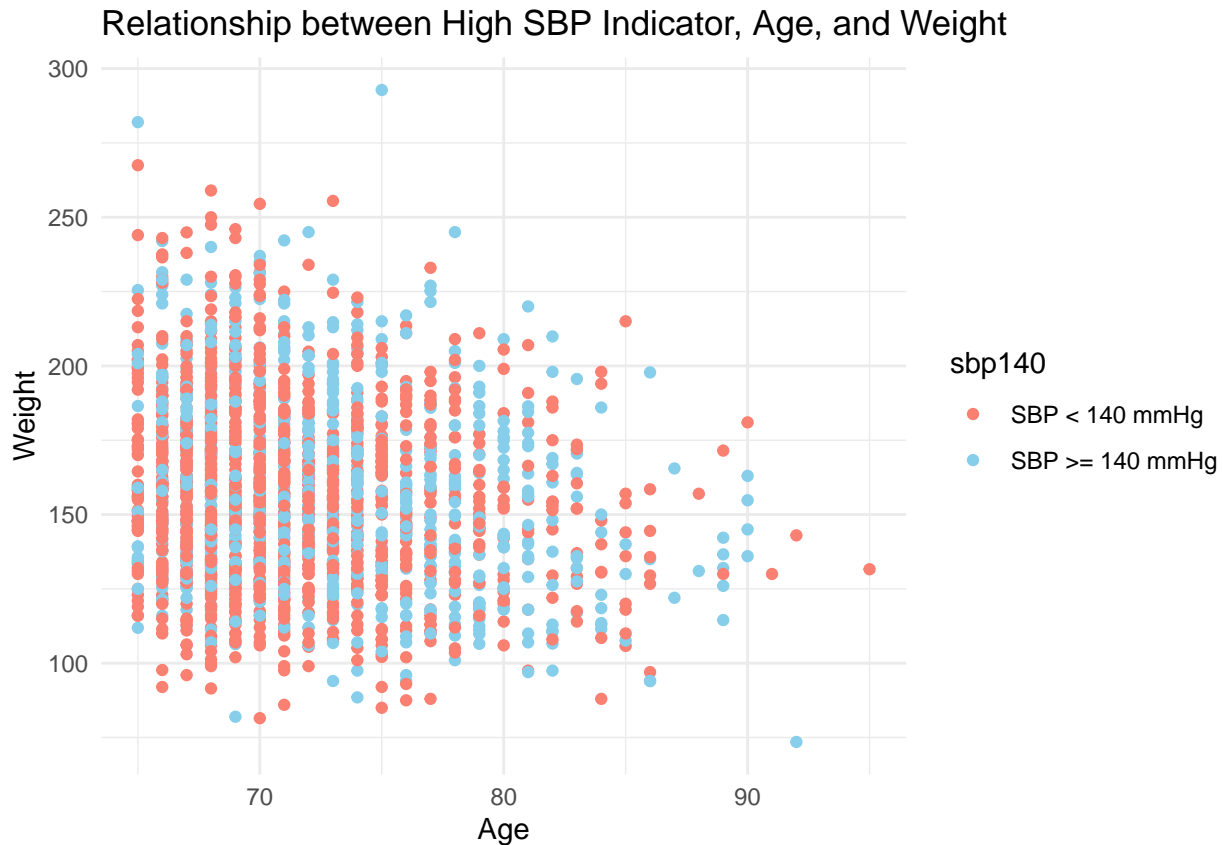
chs_clean %>%
  ggplot(aes(factor(sbp140), weight)) +
  geom_boxplot() +
  labs(title = "Relationship between High SBP Indicator and Weight")
```



(6.) Identify a Potential Confounder Confounder: Age Reason: Age is a potential confounder because it is often related to higher systolic blood pressure (sbp140) as people age. Additionally, age can influence weight due to changes in metabolism and activity levels over time.

(7.) Plot with Confounder (Age)

```
chs_clean %>%
mutate(sbp140 = factor(sbp140, levels = c(0, 1), labels = c("SBP < 140 mmHg", "SBP >= 140 mmHg"))) %>%
ggplot(aes(x = age, y = weight, color = sbp140)) +
geom_point() +
scale_color_manual(values = c("SBP < 140 mmHg" = "salmon", "SBP >= 140 mmHg" = "skyblue")) + # Custom c
theme_minimal() +
labs(
title = "Relationship between High SBP Indicator, Age, and Weight",
x = "Age",
y = "Weight"
)
```



(8.)

```
library(gtsummary)

chs_clean %>%
  select(gender, weight, age, diab, sbp140) %>%
  rename(sex = gender, diabetes = diab) %>%
  mutate(sbp140 = ifelse(sbp140 == "0", "SBP < 140 mmHg", "SBP >= 140 mmHg")) %>%
  tbl_summary(
    by = sbp140,
    digits = list(all_continuous() ~ c(2, 2)),
    statistic = all_continuous() ~ "{mean} ({sd})"
  ) %>%
  modify_header(label ~ "**Variable**") %>%
  modify_caption(caption = "Descriptive Statistics for SBP Categories")
```

Part 3: Data Analysis

(9.) Test if Proportion of High SBP is Different from 50%

```
# Hypotheses
# H0: p = 0.5 (Proportion of High SBP is equal to 50%)
# H1: p ≠ 0.5 (Proportion of High SBP is Different from 50%)

test_result <- prop.test(sum(chs_clean$sbp140 == "1"), nrow(chs_clean), p = 0.5)
test_result

##
```

Table 2: Descriptive Statistics for SBP Categories

Variable	SBP < 140 mmHg N = 1,287 ¹	SBP ≥ 140 mmHg N = 776 ¹
sex	524 (41%)	321 (41%)
weight	156.84 (30.20)	158.84 (31.34)
age	71.20 (4.70)	72.83 (5.36)
diabetes		
none	1,023 (79%)	548 (71%)
borderline	159 (12%)	114 (15%)
diabetes	105 (8.2%)	114 (15%)

¹n (%); Mean (SD)

```
## 1-sample proportions test with continuity correction
##
## data:  sum(chs_clean$sbp140 == "1") out of nrow(chs_clean), null probability 0.5
## X-squared = 126.08, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.3552568 0.3975114
## sample estimates:
##           p
## 0.3761512
```

```
# confidence interval
if (test_result$conf.int[1] <= 0.5 && test_result$conf.int[2] >= 0.5) {
  decision <- "Fail to reject H0: Proportion of High SBP is not different from 50%"
} else {
  decision <- "Reject H0: Proportion of High SBP is different from 50%"
}
decision
```

```
## [1] "Reject H0: Proportion of High SBP is different from 50%"
```

Decision and conclusion

Decision: Reject H_0 Conclusion: We have enough evidence that the proportion of high SBP is significantly different from 50%.

(10.) Test if Weight Differs by SBP Indicator

```
# Hypotheses:
# H0: Mean weight is the same for both SBP groups.
# HA: Mean weight is different for the two SBP groups.

t_test_result <- t.test(weight ~ sbp140, data = chs_clean)
t_test_result

##
## Welch Two Sample t-test
##
## data:  weight by sbp140
## t = -1.4226, df = 1586.2, p-value = 0.155
```

```
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -4.7547239 0.7571068
## sample estimates:
## mean in group 0 mean in group 1
## 156.8379 158.8367

if (t_test_result$p.value < 0.05) {
  decision <- "Reject the null hypothesis"
} else {
  decision <- "Fail to reject the null hypothesis"
}
decision

## [1] "Fail to reject the null hypothesis"
```

Decision and conclusion

Decision: Fail to reject H_0 Conclusion: We don't have enough evidence that there is a significant difference in mean weight between the two SBP groups.

(11.) Test Association Between Weight Group and SBP Indicator

```
library(dplyr)

# Creating the weight_grp variable
chs_clean <- chs_clean %>%
  mutate(weight_grp = case_when(weight < 135 ~ "< 135 lbs",
                                weight >= 135 & weight <= 160 ~ "135-160 lbs",
                                TRUE ~ "> 160 lbs")) %>%
  mutate(weight_grp = factor(weight_grp))

# (1) Hypotheses:
# H0: There is no relationship between weight group and SBP140.
# HA: There is a relationship between weight group and SBP140.

# (2) Conditions Check
# Independence: Observational units are randomly selected and each group is independent of one another.
# Expected counts: It should be 5 or more for each cell in the table

table_grp <- table(chs_clean$weight_grp, chs_clean$sbp140)
table_grp

##
##           0    1
## < 135 lbs 338 181
## > 160 lbs 558 348
## 135-160 lbs 391 247

table_grp >= 5

##
##           0    1
## < 135 lbs TRUE TRUE
## > 160 lbs TRUE TRUE
## 135-160 lbs TRUE TRUE
```

```

# (3) Perform the Chi-squared Test
test_result <- chisq.test(table_grp)
test_result

##
## Pearson's Chi-squared test
##
## data:  table_grp
## X-squared = 2.2339, df = 2, p-value = 0.3273

# (4) p-value
p_value <- test_result$p.value

# (5) Decision and Conclusion
if (p_value < 0.05) {
  decision <- "Reject the null hypothesis."
} else {
  decision <- "Fail to reject the null hypothesis."
}
decision

```

```
## [1] "Fail to reject the null hypothesis."
```

Decision: Fail to reject the H_0 . Conclusion: We don't have enough evidence that there is a significant relationship between weight group and SBP140.

Part 4: Discussion of Results

(12.)

No! While I initially hypothesized a positive association between weight and SBP, these statistical results did not provide enough evidence to support that relationship in this analysis. The t-test showed no significant difference in mean weight between SBP groups ($p = 0.155$), and the chi-squared test found no significant association between weight groups and SBP levels ($p = 0.3273$). These results indicate insufficient evidence for the hypothesized correlation.

5. I found Table of Descriptive Statistics for SBP Categories difficult but it's very important.

6. E - Excellent