**Question 3 Ideas:**

rename the 1 and other to the names of the type

using the classification labels as the X

y = number of breast cancer patients

facet grid: one graph that shows the breast cancer in each classification for each classification so there could be 7 individual graphs.

or I can do one big graph

DS HW3 q2 for facet grid

<https://www.marsja.se/r-count-the-number-of-occurrences-in-a-column-using-dplyr/> (**this is what helped me get the right answer)**

**Q3 original code:**

#Task: Create a bar chart showing the proportion of breast cancer cases and controls within each BMI category

#Note to self: referred to DS midterm

dataclean = bcdata %>%

mutate(classification = as.character(classification)) %>%

mutate(classification = recode(classification,

"1" = "Healthy Controls",

"2" = "Breast Cancer Patients")) %>%

group\_by(bmi\_category, classification) %>%

summarize(

total = n()

)

bcdata\_graphs <- dataclean %>%

ggplot(aes(x = classification, y = total, fill = classification)) +

geom\_bar(stat = "identity") +

scale\_fill\_viridis\_d() +

facet\_grid(~bmi\_category) +

labs(

title = "Proportion of Breast Cancer Cases and

Controls Within Each BMI Category",

x = "Patient Type",

y = "Distribution of Study Participants"

) + theme(axis.text.x = element\_text(angle = 90, vjust = 0.5, hjust = 1)) + theme(plot.title = element\_text(hjust = 0.5)) + theme(legend.position = "bottom")

bcdata\_graphs

**Q3: try again**

dataclean = bcdata %>%

mutate(classification = as.character(classification)) %>%

mutate(classification = recode(classification,

"1" = "Healthy Controls",

"2" = "Breast Cancer Patients")) %>%

group\_by(bmi\_category, classification) %>%

summarize(

type\_total = n(),

bmi\_cat\_total = sum(bmi\_category),

proportion = ((100\*(type\_total/bmi\_cat\_total))

))

**Question 4 helpful links:**

<https://tutorials.methodsconsultants.com/posts/logistic-regression-in-r/>

<https://data.princeton.edu/wws509/r/c3s1>

<https://www.rdocumentation.org/packages/Rcmdr/versions/2.5-3/topics/Confint>

**Question 4 First Code I tried:**

bc = bcdata %>%

mutate(classification = as.factor(classification))

logistic = bc %>%

glm(classification ~ glucose + homa + leptin + bmi + age, data = logistic, family = binomial())

summary(logistic)

logistic %>%

broom::tidy() %>%

mutate(OR = exp(estimate)) %>%

select(term, log\_OR = estimate, OR, p.value) %>%

knitr::kable(digits = 3)

**Q4 again:**

bc = bcdata %>%

mutate(classification = as.factor(classification))

logistic = glm(classification ~ glucose + homa + leptin + bmi + age, data = bc, family = binomial())

summary(logistic)

logistic %>%

broom::tidy() %>%

mutate(OR = exp(estimate)) %>%

select(term, log\_OR = estimate, OR, p.value) %>%

knitr::kable(digits = 3)

**q4 again:**

bc = bcdata %>%

mutate(classification = as.factor(classification))

logistic = glm(classification ~ glucose + homa + leptin + bmi + age, data = bc, family = binomial())

summary(logistic)

#writing function

sample = function(df) {

sample\_frac(df, replace = TRUE)

}

#from bootstrapping lecture

boot\_straps =

data\_frame(

strap\_number = 1:5000,

sample = rerun(5000, sample(bc))

)

boot\_straps

#ANALYZING

bootstraps\_results =

boot\_straps %>%

mutate(

models = map(.x = sample, ~glm(classification ~ homa, data = .x)),

results = map(models, broom::glance) #rsquare

) %>%

select(strap\_number, results) %>%

unnest(results)

bootstrap\_results %>%

group\_by(term) %>%

summarize(

ci\_lower = quantile(estimate, 0.025),

ci\_upper = quantile(estimate, 0.975))