```
#QUESTION ONE.
pnorm(1,0.7347,0.75998) #Probability of zero matches
1 - pnorm(1,0.7347, 0.75998) #Probability having at least one match a week
(1 - pnorm(1,0.7347, 0.75998))^52 # Probability of having at least one match in all 52
Weeks
#Billy statement is inaccurate given how unlikely this actually is. We can see from the
calculation that he has around a 36% chance of having at least one match a week. But when
we calculate his probability of doing this for 52 weeks, we get a very small number.

#QUESTION TWO.
library(mosaic)
library(ggplot2)
lc50 = c(16, 5, 21, 19, 10, 5, 8, 2, 7, 2, 4, 9)
M = do(2000)*mean(sample(lc50, replace = TRUE))
dim(M)
head(M)

ggplot(data = M, aes(x = M$mean)) +
  geom_histogram(color = "lightblue", fill = "lightblue", bins = 30) +
  xlab("Values of Bootstrap Mean") +
  ylab("Count") +
  ggtitle("Distribution of Bootstrap Statistic: Sample Mean") +
  geom_vline(xintercept = mean(lc50), color = "red", linetype = "dashed") +
  theme_minimal()
#QUESTION TWO, PART TWO
lc50 <- c(16, 5, 21, 19, 10, 5, 8, 2, 7, 2, 4, 9)
library(mosaic)
M <- do(2000) * mean(sample(lc50, replace = TRUE))
# Compute 95% bootstrap percentile confidence interval
lower_bound <- quantile(M$mean, 0.025)
upper_bound <- quantile(M$mean, 0.975)
cat("95% Bootstrap Percentile Confidence Interval for μLC50:", "(", lower_bound, ",",
upper_bound, ")")

#QUESTION TWO, PART THREE: Repeat your estimation of μLC50 , using the “other” confidence
interval covered in Data 602. In the context of these data, interpret the meaning of the
confidence interval. State any conditions/assumptions that are required in the computation
of this confidence interval.
lc50 = c(16, 5, 21, 19, 10, 5, 8, 2, 7, 2, 4, 9)
#density plot of the sample
densityplot(lc50, xlab="values of mean", main="Distribution of Sample Mean")
n = length(lc50)
t = qt(0.975, n-1)
UL = mean(lc50) + t*sd(lc50)/sqrt(n)
LL = mean(lc50) - t*sd(lc50)/sqrt(n)
cat("95% CI for mean lc50 measurement is", "(", LL, ",", UL, ")")
#Conditions & Assumptions Include: Using default value 0.95 confidence interval, Random
sampling, used for smaller population sizes, and unknown population standard deviation.


#QUESTION TWO, PART FOUR: Compare your results in parts (b) and (c). If you were to report
one of these confidence intervals, which would you report? Explain your answer.
#Part b results = ( 5.75 , 12.66667 )
#Part c results =  ( 4.91814 , 13.08186 )
#We want a tighter dispersion so; in this case we want to report a tighter interval so I
prefer the interval from Bootstrap

# QUESTION THREE: PART ONE
library(mosaic)
# data
n_hs_or_less <- 670
n_disagree_hs_or_less <- 348
#proportion
p_original <- n_disagree_hs_or_less / n_hs_or_less
```

```
# Bootstrap
B <- do(1000) * {
  # Sample with replacement
  sampled_data <- resample(c(rep(1, n_disagree_hs_or_less), rep(0, n_hs_or_less -
n_disagree_hs_or_less)), n_hs_or_less)
  p_bootstrap <- mean(sampled_data)

  data.frame(mean = p_bootstrap)
}

# Calculate quantiles
lower_quantile <- quantile(B$mean, 0.025)
upper_quantile <- quantile(B$mean, 0.975)

# Plot
ggplot(data = B, aes(x = mean)) +
  geom_histogram(color = "green", fill = "green") +
  xlab("Values of Bootstrap proportion") +
  ylab("Count") +
  ggtitle("Distribution of Bootstrap Statistic: Sample proportion")

#QUESTION THREE: PART TWO
# Original data
n_undergrad_or_more <- 376
n_disagree_undergrad_or_more <- 274

# Proportion
p_original_undergrad_or_more <- n_disagree_undergrad_or_more / n_undergrad_or_more

# Bootstrap
B_undergrad_or_more <- do(1000) * {
  # Sample with replacement
  sampled_data_undergrad_or_more <- resample(c(rep(1, n_disagree_undergrad_or_more),
rep(0, n_undergrad_or_more - n_disagree_undergrad_or_more)), n_undergrad_or_more)


  p_bootstrap_undergrad_or_more <- mean(sampled_data_undergrad_or_more)

  data.frame(mean = p_bootstrap_undergrad_or_more)
}

# Calculate quantiles
lower_quantile_undergrad_or_more <- quantile(B_undergrad_or_more$mean, 0.025)
upper_quantile_undergrad_or_more <- quantile(B_undergrad_or_more$mean, 0.975)

# Plot the bootstrap distribution
ggplot(data = B_undergrad_or_more, aes(x = mean)) +
  geom_histogram(color = "lightblue", fill = "lightblue") +
  xlab("Values of Bootstrap proportion") +
  ylab("Count") +
  ggtitle("Distribution of Bootstrap Statistic: Sample proportion (Undergraduate or
More)")

#QUESTION THREE PART C
# Original data for both populations
n_hs_or_less <- 670
n_disagree_hs_or_less <- 348

n_undergrad_or_more <- 376
n_disagree_undergrad_or_more <- 274

# Original proportions
p_original_hs_or_less <- n_disagree_hs_or_less / n_hs_or_less
p_original_undergrad_or_more <- n_disagree_undergrad_or_more / n_undergrad_or_more
```

```
# Bootstrap
B_hs_or_less <- do(1000) * {
  # Sample with replacement for high school or less population
  sampled_data_hs_or_less <- resample(c(rep(1, n_disagree_hs_or_less), rep(0, n_hs_or_less
- n_disagree_hs_or_less)), n_hs_or_less)


  p_bootstrap_hs_or_less <- mean(sampled_data_hs_or_less)

  sampled_data_undergrad_or_more <- resample(c(rep(1, n_disagree_undergrad_or_more),
rep(0, n_undergrad_or_more - n_disagree_undergrad_or_more)), n_undergrad_or_more)

  p_bootstrap_undergrad_or_more <- mean(sampled_data_undergrad_or_more)

  diff_proportions <- p_bootstrap_undergrad_or_more - p_bootstrap_hs_or_less

  data.frame(diff_proportions = diff_proportions)
}

# Solution
lower_quantile_diff_proportions <- quantile(B_hs_or_less$diff_proportions, 0.025)
upper_quantile_diff_proportions <- quantile(B_hs_or_less$diff_proportions, 0.975)

# Plot the bootstrap distribution
ggplot(data = B_hs_or_less, aes(x = diff_proportions)) +
  geom_histogram(color = "lightblue", fill = "lightblue") +
  xlab("Difference in Proportions (Uni - HS)") +
  ylab("Count") +
  ggtitle("Distribution of Bootstrap Statistic: Difference in Proportions") +
  geom_vline(xintercept = lower_quantile_diff_proportions, color = "red") +
  geom_vline(xintercept = upper_quantile_diff_proportions, color = "red")

#QUESTION THREE, PART FOUR: # Calculate the 95% bootstrap percentile confidence interval
conf_interval_diff_proportions <- quantile(B_hs_or_less$diff_proportions, c(0.025, 0.975))
conf_interval_diff_proportions
#Findings from bootstrap simulation show that the proportion of individuals with an
education level higher than high school who disagree with the statement about the clarity
of the science around vaccination is greater than the proportion of individuals with an
education level equal to or less than high school. This is supported by the distribution
of simulated differences, which yielded values greater than 0. The 95% confidence interval
for the difference in proportions, ranging from 0.146708 to 0.2651816, further supports
this. This interval suggests that, with 95% confidence, the true difference in proportions
favors individuals with a higher education level than high school.

#QUESTION FOUR PART A
prop.test(163 + 2, 1000 + 4, correct=FALSE)$conf

#QUESTION FOUR PART B
trials <- 1000
pboot.inf <- 163 / trials
inf <- rbinom(trials, 1000, pboot.inf) / trials
quantile(inf, c(0.025, 0.975))

#QUESTION FOUR PART C
#Yes, we can infer this - the proportion of Canadians who believe inflation is the most
important national issue has increased. We can infer this because 0.13 falls out of the
lower bound of our previous calculation.


#QUESTION FIVE, PART ONE
prop.test(128+2, 399+4, correct=FALSE)$conf

#QUESTION FIVE, PART TWO
conservative = 128 + 2  # Conservative Respondants
n_total = 399 + 4  # Total
```

```
num_replications = 1000
bootstrap <- numeric(num_replications)

for (i in 1:num_replications) {

  bootstrap_sample <- sample(c(rep(1, conservative), rep(0, n_total - conservative)),
n_total, replace = TRUE)


  bootstrap[i] <- sum(bootstrap_sample) / n_total
}
head(bootstrap)

#QUESTION FIVE, PART THREE
conf_interval <- quantile(bootstrap, c(0.025, 0.975))
conf_interval

#QUESTION FIVE, PART FOUR
#Would want to use the tighter dispersion, for more accuracy. In this case would use
bootstrap.
```