00:02:49 Sam Holland: Hi Kelsey, are we doing 3 or 310 today?
00:03:23 Kelsey MacCuish: 3:10!
00:03:29 Lian Hsiao: I was wondering if we'll be going over each problem together?
00:03:35 Sam Holland: Ok awesome thank you
00:03:37 Kelsey MacCuish: we'll always start 10 after :-)
00:03:49 Kelsey MacCuish: we'll try to go over as much as we can together!
00:04:20 Kelsey MacCuish: today will follow a similar format as last week - will poll you all again next week to see if you want to follow a different type of lab structre
00:04:24 Kelsey MacCuish: structure
00:05:39 Julia Hankin: Hey Kelsey! :) Can you remind me what material will be included on Quiz 1? Is it primarily content from last week (using dplyr for example)?
00:07:04 Kelsey MacCuish: Hi Julia (and all!) yes - quiz 1 material will be from last wednesdays lect, friday's lect, and monday's lect
00:07:23 Kelsey MacCuish: will always follow the format of previous week's wed/fri and concurrent week's monday
00:08:18 Kelsey MacCuish: so yes a lot of dplyr (since that was a lot of friday and monday) but don't forget to look at the very first lect! (wed 8/25(
00:08:33 Julia Hankin: Awesome, thank you!
00:08:53 Julia Hankin: Also, very much grooving to the soundtrack rn :)
00:09:03 Kelsey MacCuish: :-)))
00:10:42 Lian Hsiao: What's the structure of the quiz like? Is it multiple choice or we type in our own code? And how many questions/ how long is it suppose to take?
00:11:44 Kelsey MacCuish: the quiz is all types of qs - its meant to take about 20 mins but you have an hour
00:12:32 Lian Hsiao: Got it, thank you!
00:13:47 Ijeoma Uche: When will it be released?
00:15:26 Ijeoma Uche: Will it automatically be graded?
00:15:39 Maddy Griffith: If there's time today, could you talk about the grading for lab1? Was a little confusing from the piazza chat
00:16:11 Kelsey MacCuish: no quizzes are not automatically graded
00:17:52 Kelsey MacCuish: 1. when do you use a bar plot vs. histogram?
00:18:04 Kelsey MacCuish: 2. are bar plots and histograms used to visualize bivariate or univariate data?
00:18:13 Kelsey MacCuish: (and what is bivariate/univaraite data)
00:18:22 Gabriela Gonzalez: I'm unfamiliar with those types of data
00:18:24 Chitra Nambiar: univariate
00:20:10 Alexandra Potter: can't you have a bar chart of bivariate data?
00:20:16 Ximena Verduzco-Villanueva: Can yo repeat the example of univariate please?
00:20:18 Kelsey MacCuish: 3. you have a dataset flu_data.csv
00:20:29 Kelsey MacCuish: flu_data

00:21:38 Olufunke Fasawe: Where do we find the flu_data.csv file?
00:23:55 Sam Holland: flu_data <- read.csv(flu_data.csv) ? I believe you just said to read it in, correct?
00:24:38 Ximena Verduzco-Villanueva: yes
00:30:47 Taylor Yoo: sounds good to me :)
00:30:48 Nora Povejsil: I want to work on the lab! Thank you
00:30:52 Maddy Griffith: Sounds good
00:30:54 Lupita Ambriz: Lab please !
00:30:54 Silvana Larrea: Me too
00:30:54 Gabriela Gonzalez: lab please!
00:36:27 Olufunke Fasawe: What is the code for clearing the Environment?
00:36:43 Maddy Griffith: You can use the little broom icon!
00:38:03 Ximena Verduzco-Villanueva: I don’t see my lab 02 on my file folder.
00:38:36 Ijeoma Uche: I get this error: “28 Aug 2021 02:05:36 [rsession-rstudio] ERROR session hadabend; LOGGED FROM: rstudio::core::Error {anonymous}::rInit(const rstudio::r::session::RInitInfo&) src/cpp/session/SessionMain.cpp:655”
00:38:40 Ijeoma Uche: Is this normal?
00:39:50 Genesis Navarrete: I got that error too
00:42:25 Andriana Marijic Buljubasic (she/her): I just use the broom
00:42:32 Lian Hsiao: It’s rm(list=ls()) also
00:45:13 Nora Povejsil: I have the same issue
00:53:17 Julia Hankin: Not sure if you’re planning to review the different functions to get to know the data set, but I’m still confused by the str() function in particular. Any tips to understanding how this function helps us?
00:54:25 Kelsey MacCuish: reviewing ws noq~
00:54:26 Kelsey MacCuish: now!
00:54:32 Ximena Verduzco-Villanueva: Thats how mine is showing up @Nora
00:55:14 Ijeoma Uche: gives us a compacted version of all the variables/ data
00:55:25 Rachel Harvill: gives you an idea of the structure of the dataset
00:56:07 Sam Holland: Are the columns considered strings? So str summarizes strings?
00:56:21 Ijeoma Uche: states the names of the categories/ variables
00:56:32 Ijeoma Uche: shows us the first part of the data set/ table frame or matrix
00:56:52 Yichao Sun: Show the dimension of the function, the result will display the number of rows by the number of columns.
00:57:17 Ijeoma Uche: 8?
00:57:20 Silvana Larrea: 137
00:57:23 Annalisa Watson: 137
00:58:32 Maddy Griffith: countries
00:58:32 Taylor Yoo: countries
00:58:33 Silvana Larrea: countires

```
01:00:05 Ijeoma Uche:     CS_rate<- CS_data_raw  %>%
  mutate(CS_rate_100 = CS_rate * 100)
CS_rate
01:02:02 Gabriela Gonzalez:       I am told CS_rate was not found
01:02:07 Gabriela Gonzalez:       as an error
01:02:41 Gabriela Gonzalez:       i fixed it nvm
01:07:28 Mariah Jiles (she/her):  income group?
01:07:44 Yichao Sun:      Hi Kelsey
01:07:45 Yichao Sun:      i got the following error code
01:07:46 Yichao Sun:      ── Error (q3.R:19:3): p3a
──────────────────────────────────────────────────────────
Error: object 'CS_data' not found
Backtrace:
 1. testthat::expect_true(is.data.frame(CS_data)) tests/q3.R:19:2
 4. base::is.data.frame(CS_data)

── Error (q3.R:24:3): p3b
──────────────────────────────────────────────────────────
Error: object 'CS_data' not found
Backtrace:
 1. testthat::expect_true(...) tests/q3.R:24:2
 4. base::nrow(CS_data)

── Error (q3.R:29:3): p3c
──────────────────────────────────────────────────────────
Error: object 'CS_data' not found
Backtrace:
 1. testthat::expect_true(all.equal(CS_data$CS_rate[1] * 100,
CS_data$CS_rate_100[1])) tests/q3.R:29:2
 4. base::all.equal(CS_data$CS_rate[1] * 100, CS_data$CS_rate_100[1])

Test p3a failed:
    Error: object 'CS_data' not found

Test p3b failed:
    Error: object 'CS_data' not found

Test p3c failed:
    Error: object 'CS_data' not found
01:07:46 Chitra Nambiar:  Income group / income group order
01:07:52 Silvana Larrea:  Region
01:07:55 Stacy (Seohyun) Ahn:     Region
01:08:00 Mariah Jiles (she/her):  region
01:10:43 Olufunke Fasawe: Going well so far
01:18:05 Maddy Griffith:  p6 <- ggplot(data=CS_data,
aes(x=Income_Group)) + geom_bar()
01:18:09 Rachel Harvill:  ggplot(CS_data, aes(x=Income_Group)) +
geom_bar()
01:18:14 Gabriela Gonzalez:       I put p6 <- ggplot(CS_data, aes (x=
region, y = percent_Income_group)) +
```

```
  geom_bar(stat = "identity") +
  labs(y = "Percent", x = "") +
  theme_minimal(base_size = 15)
01:18:21 Taylor Yoo:      p6 <- ggplot(CS_data, aes(x = Income_Group))
+
  geom_bar() +
  labs(x="Income Group")
01:21:22 Olufunke Fasawe: How do you print p6?
01:21:46 Sam Holland:     I run the exact code you dopy and pasted but
it says "list of 9" in my environment...
01:22:02 Sam Holland:     It doesn't show a bar chart
01:22:45 Stacy (Seohyun) Ahn:     p7 <- ggplot(data=CS_data,
aes(x=Income_Group_order)) +
  geom_bar() +
  labs (x="Income_Goupr_order")
p7
01:23:05 Stacy (Seohyun) Ahn:     yayy
01:25:09 Mariah Jiles (she/her):  Income group order
01:25:11 Stacy (Seohyun) Ahn:     income group order
01:25:15 Silvana Larrea:  The ordered variable
01:25:22 Rachel Harvill:  I preferred ordered, it was easier to look at
01:26:36 Anai Ramos:      It wasn’t aggregated data
01:26:36 Olufunke Fasawe: We did not use it because there is no data on
the y axis
01:26:37 Chitra Nambiar:  not used because we want R to do the
aggregation
01:26:54 Ijeoma Uche:     No because we did not used aggregative data ?
01:28:24 Nora Povejsil:   It works now! Thanks
01:32:21 Anai Ramos:      Yes
01:32:28 Chitra Nambiar:  yes
01:42:17 Stacy (Seohyun) Ahn:     Also, are we supposed to see our
grades from lab 1? Mine still says ungraded.
01:42:37 Stacy (Seohyun) Ahn:     oh just checked again
01:42:40 Stacy (Seohyun) Ahn:     it’s there
01:42:42 Stacy (Seohyun) Ahn:     thanks Kelsey!!
01:42:56 Ijeoma Uche:     p10 <- ggplot(CS_data, aes(x = Region)) +
  geom_bar() +
  labs(x="Region", y= "number of countries")
p10
01:43:33 Jessica Wright:  it’s labeled Region in the data I think
01:44:15 Hiruni Jayasekera (she/her):     when I go to look at my
plots they’re “list of 9” instead of a plot
01:44:25 Ijeoma Uche:     Europe & Central Asia
01:44:33 Ijeoma Uche:     North America
01:44:33 Silvana Larrea:  North America
01:44:34 Maddy Griffith:  N. America
01:51:05 Chitra Nambiar:  ggplot(dat = CS_data, aes(x = GDP_2006)) +
geom_histogram(binwidth = 5000, col = "white", fill = "forest green")
01:51:12 Yichao Sun:      ggplot(CS_data, aes(x = GDP_2006)) +
geom_histogram()
```

01:51:14 Julia Hankin: p13 <- ggplot(CS_data, aes(x = GDP_2006)) +
  geom_histogram() +
  labs(x="2006 GDP")
p13
01:51:18 Ijeoma Uche: p13 <-ggplot(data= CS_data, aes(x
=GDP_2006)) +
  geom_histogram(col= "black", binwidth = 8000) +
  labs(x= "GDP_2006", y="count")+
  theme_minimal(base_size = 15)
p13
01:51:24 Maddy Griffith: p13 <- ggplot(data = CS_data,
aes(x=GDP_2006)) + geom_histogram()
p13
01:54:42 Jessica Wright: do you have to add data= C_data? I ran it
without typing “data=“ and just wrote “CS_data” and it worked
01:56:52 Sam Holland: Can I ask what exactly changing binwidth
does?
01:56:59 Sam Holland: like data-wise
01:59:50 Christopher Patterson: This has been very helpful !
02:00:12 Stacy (Seohyun) Ahn: super helpful!
02:00:16 Gabriela Gonzalez: this has been helpful
02:00:27 Nora Povejsil: What does Error: `quo` must be a quosure
mean?
02:00:31 Julia Hankin: I feel better about ggplot than I did
before :)
02:00:44 Rachel Harvill: Kelsey I have a question about submitting the
lab to grade scope, the code currently uses lab02.rmd, not
lab02_UPDATED.rmd
02:01:04 Rachel Harvill: Oh you are covering it :)
02:02:39 Silvana Larrea: Kelsey sorry, can you repeat the renaming the
file part?
02:03:33 Silvana Larrea: Thanks!
02:04:21 Ijeoma Uche: Can we go owe 17? Do we need to rename the
graphs?
02:04:27 Ijeoma Uche: over*
02:04:28 Stacy (Seohyun) Ahn: ^
02:04:56 Gabriela Gonzalez: questions
02:04:59 Yichao Sun: Questions
02:05:01 Olufunke Fasawe: Questions
02:06:48 Ijeoma Uche: 5?
02:07:01 Olufunke Fasawe: I used 3
02:11:19 Ijeoma Uche: p16 <-ggplot(data= CS_data, aes(x
=CS_rate_100)) +
  geom_histogram(col= "white", binwidth = 5, fill = "light blue") +
  labs(x= "CS_rate_100", y="count", title = "Cesarean Delivery Rates")
+
  theme_minimal(base_size = 10)
p16
02:11:20 Gabriela Gonzalez: p16 <- ggplot(data = CS_data, aes(x
= CS_rate_100)) + geom_histogram(col = "white", binwidth = 5) +

```
  labs(x= "CS Rate")
p16
```
02:11:26 Olufunke Fasawe: `ggplot(CS_data, aes(x = CS_rate_100)) + geom_histogram(col = "blue", binwidth = 4) + labs(x = "CS_rate_100") + theme_minimal(base_size = 15)`
02:14:08 Olufunke Fasawe: How do you add a title to the chart?
02:15:02 Aliza Adler: Sorry, to reiterate, we only have to “polish” one plot in Q16, right? We don’t have to go back and add labels, etc. to all of the previous plots?
02:15:30 Aliza Adler: Thanks!
02:16:23 Gabriela Gonzalez: i’m confused about #17
02:17:20 Sam Holland: `ggsave("CS_plot.png", plot=last_plot())`
02:17:26 Sam Holland: It saved a 7x7 image somehow
02:18:18 Ijeoma Uche: `ggplot(CS_data, aes(CS_rate_100, wt)) + geom_point()`
`ggsave("Cesarean Delivery Rates.pdf")` I don’t know what wt means
02:22:51 Ijeoma Uche: Thank you!This was helpful
02:24:23 Chitra Nambiar: thank you!
02:25:23 Mariah Jiles (she/her): Do you mind going over question 17 again?
02:25:37 Aliza Adler: Under the submission instructions, could you explain what #1 means? “1. Any image files you add that are needed to knit the file are in the `src` folder and file paths are specified accordingly. “
02:27:16 Olufunke Fasawe: my image is not coming out
02:27:23 Aliza Adler: Okay got it! Thanks!!
02:27:36 Olufunke Fasawe: I used this code : `ggsave("CS_hist.png", plot = p16, device = png, scale = 1, units = "in", width = 10, height = 10)`
02:27:46 Ijeoma Uche: What google form do we need to complete for this week?