

Question 1

The first function works because the double curly brackets inside the parentheses ({{}}) will treat column as placeholder for any column name. In this case, we use put "type" as a column name in the group_and_tally function. The second function will not work because the parentheses () will treat "column" as a literal column name itself, meaning that the column name we want to group by (type) won't be evaluated properly.

Question 2

```
git clone git@github.com:collings512/BIOS512_Collin_Stewart.git
```

Question 3

```
ls -al ~/.ssh
ssh-keygen -t ed25519 -C "cgstew@unc.edu"
eval "$(ssh-agent -s)"
ssh-add ~/.ssh/id_ed25519
cat ~/.ssh/id_ed25519.pub
ssh -T git@github.com
```

Question 4

```
4a.) cd ~/BIOS512/BIOS512_Collin_Stewart
mkdir HW2
echo "This is for homework 2." > HW2/HW2.md
ls HW2
cat HW2/HW2.md
```

```
4b.) git add HW2/HW2.md
git status
```

This shows:
On branch master
Your branch is up to date with 'origin/master'.

Changes to be committed:
(use "git restore --staged ..." to unstage)
new file: HW2/HW2.md

```
4c.) git commit -m "Add HW2.md with initial text"
```

[master 5086884] Add HW2.md with initial text

1 file changed, 1 insertion(+)

create mode 100644 HW2/HW2.md

git log --oneline

5086884 (HEAD -> master) Add HW2.md with initial text

27b7c1e (origin/master, origin/HEAD) Adding data for homework 3

2b1886d Create Extra.tips

a6eed2f Create apt.txt

5361c6d Create install.R

c612d52 Create runtime.txt

1761634 Adding starter HW3 file

e751db2 Create HW3 folder

8eb3bbf Delete Homework_3

30419ce Create Homework_3

a9a6ae8 Updated file where each question has its own cell

4ebdbb5 Add my .ipynb file of HW 2

2944ebf Create placeholder.txt

a67aa7f Create placeholder.txt

244dabc initial commit

4d.) echo "# Homework 2" | cat - HW2/HW2.md > HW2/temp && mv HW2/temp

HW2/HW2.md

cat HW2/HW2.md

Returns:

Homework 2

This is for homework 2.

4e.) git diff

Returns:

warning: in the working copy of 'HW2/HW2.md', LF will be replaced by CRLF the next time

Git touches it

diff --git a/HW2/HW2.md b/HW2/HW2.md

index 1a010d3..e202d7f 100644

--- a/HW2/HW2.md

+++ b/HW2/HW2.md

@@ -1 +1,2 @@

+# Homework 2

This is for homework 2.

4f.) git status

Returns:

On branch master

Your branch is ahead of 'origin/master' by 1 commit.

(use "git push" to publish your local commits)

Changes not staged for commit:

(use "git add ..." to update what will be committed)

(use "git restore ..." to discard changes in working directory)

modified: HW2/HW2.md

no changes added to commit (use "git add" and/or "git commit -a")

4g.) git add HW2/HW2.md

warning: in the working copy of 'HW2/HW2.md', LF will be replaced by CRLF the next time Git touches it

git commit -m "Add title to HW2.md"

[master 46f48ce] Add title to HW2.md

1 file changed, 1 insertion(+)

git log --oneline

46f48ce (HEAD -> master) Add title to HW2.md

5086884 Add HW2.md with initial text

27b7c1e (origin/master, origin/HEAD) Adding data for homework 3

2b1886d Create Extra.tips

a6eed2f Create apt.txt

5361c6d Create install.R

c612d52 Create runtime.txt

1761634 Adding starter HW3 file

e751db2 Create HW3 folder

8eb3bbf Delete Homework_3

30419ce Create Homework_3

a9a6ae8 Updated file where each question has its own cell

4ebdbb5 Add my .ipynb file of HW 2

2944ebf Create placeholder.txt

a67aa7f Create placeholder.txt

244dabc initial commit

4h.) git help

git help log

git help show

In the documentation:

```
git log --since="2 weeks ago" -- gitk
```

Show the changes during the last two weeks to the file gitk. The -- is necessary to avoid confusion with the branch named gitk

```
In [2]: library(tidyverse)
if (!dir.exists("intermediate")) dir.create("intermediate", recursive = TRUE)
if (!exists("mdpre")) mdpre <- function(x) { print(x) }
if (!exists("ggmd")) ggmd <- function(p) { print(p) }
```

```
— Attaching core tidyverse packages — tidyverse 2.0.0 —
✓ dplyr      1.1.2      ✓ readr      2.1.4
✓ forcats    1.0.0      ✓ stringr    1.5.0
✓ ggplot2    3.4.2      ✓ tibble     3.2.1
✓ lubridate  1.9.2      ✓ tidyr      1.3.0
✓ purrr      1.0.1

— Conflicts — tidyverse_conflicts() —
✗ dplyr::filter() masks stats::filter()
✗ dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

Question 5

```
In [4]: getwd()

pat_names <- read.csv("Homework_3/patient_names.csv")
pat_props <- read.csv("Homework_3/patient_properties.csv")

pat_names1 <- pat_names %>%
  mutate(
    BIRTHDATE = as.Date(BIRTHDATE, format = "%m/%d/%y"),
    DEATHDATE = as.Date(DEATHDATE, format = "%m/%d/%y")
  )

head(pat_names1, 10)
head(pat_props, 10)
```

"/home/jovyan"

A data.frame: 10 × 7

	ID	BIRTHDATE	DEATHDATE	FIRST	LAST	CITY	
	<chr>	<date>	<date>	<chr>	<chr>	<chr>	
1	5605b66b-e92d-c16c-1b83-b8bf7040d51f	1977-03-19	NA	Nikita578	Erdman779	Quincy	Massach
2	6e5ae27c-8038-7988-e2c0-25a103f01bfa	2040-02-19	NA	Zane918	Hodkiewicz467	Boston	Massach
3	8123d076-0886-9007-e956-d5864aa121a7	2058-06-04	NA	Quinn173	Marquardt819	Quincy	Massach
4	770518e4-6133-648e-60c9-071eb2f0e2ce	2028-12-25	2017-09-29	Abel832	Smitham825	Boston	Massach
5	f96addf5-81b9-0aab-7855-d208d3d352c5	2028-12-25	2014-02-23	Edwin773	Labadie908	Boston	Massach
6	8e9650d1-788a-78f9-4a28-d08f7f95354a	2028-12-25	NA	Frankie174	Oberbrunner298	Boston	Massach
7	183df435-4190-060e-8f8e-bf63c572b266	2057-11-08	NA	Eilene124	Walsh511	Cambridge	Massach
8	720560d4-51da-c38c-ee90-c15935278df1	1972-06-27	NA	Lowell343	Price929	Quincy	Massach
9	217851b0-5f47-d376-18b9-0fe4ba77207e	2054-03-06	NA	Adrian111	Gleason633	Boston	Massach
10	ff331e5c-ab16-e218-f39a-63e11de1ed75	2027-07-10	NA	Eugene421	Abernathy524	Boston	Massach



A data.frame: 10 × 3

	ID	property	value
	<chr>	<chr>	<chr>
1	5605b66b-e92d-c16c-1b83-b8bf7040d51f	MARITAL	M
2	5605b66b-e92d-c16c-1b83-b8bf7040d51f	RACE	white
3	5605b66b-e92d-c16c-1b83-b8bf7040d51f	ETHNICITY	nonhispanic
4	5605b66b-e92d-c16c-1b83-b8bf7040d51f	GENDER	F
5	6e5ae27c-8038-7988-e2c0-25a103f01bfa	MARITAL	M
6	6e5ae27c-8038-7988-e2c0-25a103f01bfa	RACE	white
7	6e5ae27c-8038-7988-e2c0-25a103f01bfa	ETHNICITY	nonhispanic
8	6e5ae27c-8038-7988-e2c0-25a103f01bfa	GENDER	M
9	8123d076-0886-9007-e956-d5864aa121a7	MARITAL	M
10	8123d076-0886-9007-e956-d5864aa121a7	RACE	white

Question 6

```
In [5]: library(dplyr)
library(tidyr)

pat_props_wide <- pat_props %>%
  pivot_wider(
    names_from = property,
    values_from = value
  )

head(pat_props_wide,10)
```

A tibble: 10 × 5

	ID	MARITAL	RACE	ETHNICITY	GENDER
	<chr>	<chr>	<chr>	<chr>	<chr>
5605b66b-e92d-c16c-1b83-b8bf7040d51f		M	white	nonhispanic	F
6e5ae27c-8038-7988-e2c0-25a103f01bfa		M	white	nonhispanic	M
8123d076-0886-9007-e956-d5864aa121a7		M	white	nonhispanic	M
770518e4-6133-648e-60c9-071eb2f0e2ce		M	white	hispanic	M
f96addf5-81b9-0aab-7855-d208d3d352c5		M	white	hispanic	M
8e9650d1-788a-78f9-4a28-d08f7f95354a		M	white	hispanic	M
183df435-4190-060e-8f8e-bf63c572b266		M	asian	nonhispanic	F
720560d4-51da-c38c-ee90-c15935278df1		M	white	nonhispanic	M
217851b0-5f47-d376-18b9-0fe4ba77207e		S	black	hispanic	M
ff331e5c-ab16-e218-f39a-63e11de1ed75		M	native	hispanic	M

Question 7

```
In [6]: pat_join <- pat_names1 %>%
  left_join(
    pat_props_wide,
    by = "ID"
  )

head(pat_join, 10)
```

A data.frame: 10 × 11

	ID	BIRTHDATE	DEATHDATE	FIRST	LAST	CITY	
	<chr>	<date>	<date>	<chr>	<chr>	<chr>	
1	5605b66b-e92d-c16c-1b83-b8bf7040d51f	1977-03-19	NA	Nikita578	Erdman779	Quincy	Massach
2	6e5ae27c-8038-7988-e2c0-25a103f01bfa	2040-02-19	NA	Zane918	Hodkiewicz467	Boston	Massach
3	8123d076-0886-9007-e956-d5864aa121a7	2058-06-04	NA	Quinn173	Marquardt819	Quincy	Massach
4	770518e4-6133-648e-60c9-071eb2f0e2ce	2028-12-25	2017-09-29	Abel832	Smitham825	Boston	Massach
5	f96addf5-81b9-0aab-7855-d208d3d352c5	2028-12-25	2014-02-23	Edwin773	Labadie908	Boston	Massach
6	8e9650d1-788a-78f9-4a28-d08f7f95354a	2028-12-25	NA	Frankie174	Oberbrunner298	Boston	Massach
7	183df435-4190-060e-8f8e-bf63c572b266	2057-11-08	NA	Eilene124	Walsh511	Cambridge	Massach
8	720560d4-51da-c38c-ee90-c15935278df1	1972-06-27	NA	Lowell343	Price929	Quincy	Massach
9	217851b0-5f47-d376-18b9-0fe4ba77207e	2054-03-06	NA	Adrian111	Gleason633	Boston	Massach
10	ff331e5c-ab16-e218-f39a-63e11de1ed75	2027-07-10	NA	Eugene421	Abernathy524	Boston	Massach

Question 8


```
In [7]: #Each first and last name has 3 digits stuck on the end. They need to be trimmed  
  
pat_join$FIRST <- gsub("[0-9]+$", "", pat_join$FIRST)  
pat_join$LAST  <- gsub("[0-9]+$", "", pat_join$LAST)  
  
head(pat_join,10)
```

A data.frame: 10 × 11

	ID	BIRTHDATE	DEATHDATE	FIRST	LAST	CITY	STATE
	<chr>	<date>	<date>	<chr>	<chr>	<chr>	<chr>
1	5605b66b-e92d-c16c-1b83-b8bf7040d51f	1977-03-19	NA	Nikita	Erdman	Quincy	Massachusetts
2	6e5ae27c-8038-7988-e2c0-25a103f01bfa	2040-02-19	NA	Zane	Hodkiewicz	Boston	Massachusetts
3	8123d076-0886-9007-e956-d5864aa121a7	2058-06-04	NA	Quinn	Marquardt	Quincy	Massachusetts
4	770518e4-6133-648e-60c9-071eb2f0e2ce	2028-12-25	2017-09-29	Abel	Smitham	Boston	Massachusetts
5	f96addf5-81b9-0aab-7855-d208d3d352c5	2028-12-25	2014-02-23	Edwin	Labadie	Boston	Massachusetts
6	8e9650d1-788a-78f9-4a28-d08f7f95354a	2028-12-25	NA	Frankie	Oberbrunner	Boston	Massachusetts
7	183df435-4190-060e-8f8e-bf63c572b266	2057-11-08	NA	Eilene	Walsh	Cambridge	Massachusetts
8	720560d4-51da-c38c-ee90-c15935278df1	1972-06-27	NA	Lowell	Price	Quincy	Massachusetts
9	217851b0-5f47-d376-18b9-0fe4ba77207e	2054-03-06	NA	Adrian	Gleason	Boston	Massachusetts
10	ff331e5c-ab16-e218-f39a-63e11de1ed75	2027-07-10	NA	Eugene	Abernathy	Boston	Massachusetts



Question 9

```
In [8]: p_cat_cols <- setdiff(names(pat_join), c("ID", "FIRST", "LAST"))
  for (var in p_cat_cols) {
    if(is.character(pat_join[[var]]) || is.factor(pat_join[[var]]) || inherits(pat_
      counts <- pat_join %>%
      count(.data[[var]]) %>%
      arrange(desc(n))
      cat("\n###", var, "\n")
      mdpre(counts)
    }
  }
```

```

#### BIRTHDATE
      BIRTHDATE n
1    2025-11-17 4
2    2027-04-22 4
3    2029-01-28 4
4    2029-05-15 4
5    2036-01-12 4
6    2025-08-10 3
7    2025-11-16 3
8    2026-08-07 3
9    2027-02-05 3
10   2028-11-04 3
11   2028-12-25 3
12   2030-11-03 3
13   2033-08-24 3
14   1972-04-25 2
15   1973-10-28 2
16   1973-11-17 2
17   1976-12-11 2
18   1977-01-10 2
19   1982-08-09 2
20   1984-06-14 2
21   2022-05-31 2
22   2023-09-22 2
23   2025-01-15 2
24   2025-02-09 2
25   2025-05-10 2
26   2025-06-01 2
27   2025-11-20 2
28   2026-07-14 2
29   2026-12-27 2
30   2027-06-04 2
31   2027-09-08 2
32   2027-10-16 2
33   2029-06-30 2
34   2029-10-14 2
35   2030-04-25 2
36   2030-12-14 2
37   2031-02-27 2
38   2031-05-04 2
39   2032-09-12 2
40   2032-11-06 2
41   2033-05-23 2
42   2033-08-13 2
43   2033-09-02 2
44   2034-10-28 2
45   2035-02-15 2
46   2035-04-16 2
47   2035-06-25 2
48   2036-02-07 2
49   2036-06-11 2
50   2036-06-26 2
51   2036-11-08 2
52   2037-09-04 2
53   2038-02-26 2
54   2039-07-05 2

```

55	2039-09-23	2
56	2041-12-18	2
57	2042-05-18	2
58	2042-06-18	2
59	2042-08-03	2
60	2046-01-09	2
61	2046-08-05	2
62	2048-03-05	2
63	2050-12-23	2
64	2051-03-01	2
65	2051-05-03	2
66	2052-04-01	2
67	2052-11-02	2
68	2054-01-14	2
69	2054-08-26	2
70	2057-01-19	2
71	2058-06-29	2
72	2060-09-13	2
73	2063-09-24	2
74	2064-01-05	2
75	2064-07-01	2
76	2066-11-09	2
77	1969-01-11	1
78	1969-01-26	1
79	1969-02-10	1
80	1969-02-14	1
81	1969-02-17	1
82	1969-03-21	1
83	1969-05-06	1
84	1969-05-15	1
85	1969-07-12	1
86	1969-11-20	1
87	1969-12-27	1
88	1969-12-30	1
89	1970-04-08	1
90	1970-04-16	1
91	1970-06-03	1
92	1970-09-08	1
93	1970-09-17	1
94	1970-11-26	1
95	1970-12-02	1
96	1971-01-26	1
97	1971-01-30	1
98	1971-03-31	1
99	1971-05-10	1
100	1971-05-16	1
101	1971-06-19	1
102	1971-07-05	1
103	1971-07-08	1
104	1971-08-19	1
105	1971-08-26	1
106	1971-10-27	1
107	1971-11-06	1
108	1972-01-15	1
109	1972-01-25	1
110	1972-04-15	1

111 1972-06-10 1
112 1972-06-27 1
113 1972-06-29 1
114 1972-07-30 1
115 1972-10-24 1
116 1973-01-06 1
117 1973-01-24 1
118 1973-02-16 1
119 1973-02-26 1
120 1973-03-12 1
121 1973-04-19 1
122 1973-05-16 1
123 1973-05-25 1
124 1973-06-12 1
125 1973-08-28 1
126 1973-09-01 1
127 1973-09-09 1
128 1974-01-20 1
129 1974-01-30 1
130 1974-04-21 1
131 1974-04-22 1
132 1974-04-27 1
133 1974-05-12 1
134 1974-06-19 1
135 1974-07-09 1
136 1974-07-15 1
137 1974-07-29 1
138 1974-10-03 1
139 1974-12-06 1
140 1974-12-22 1
141 1975-01-24 1
142 1975-02-05 1
143 1975-02-22 1
144 1975-03-13 1
145 1975-04-29 1
146 1975-05-07 1
147 1975-05-20 1
148 1975-05-31 1
149 1975-06-27 1
150 1975-07-27 1
151 1975-09-12 1
152 1975-09-19 1
153 1975-11-14 1
154 1975-12-06 1
155 1975-12-30 1
156 1976-01-07 1
157 1976-01-22 1
158 1976-02-27 1
159 1976-04-04 1
160 1976-04-10 1
161 1976-05-18 1
162 1976-05-26 1
163 1976-08-01 1
164 1976-08-18 1
165 1976-09-02 1
166 1976-10-13 1

167 1976-10-18 1
168 1976-10-31 1
169 1976-12-01 1
170 1977-01-02 1
171 1977-01-30 1
172 1977-02-05 1
173 1977-02-08 1
174 1977-03-19 1
175 1977-04-05 1
176 1977-04-19 1
177 1977-04-29 1
178 1977-05-16 1
179 1977-05-28 1
180 1977-06-29 1
181 1977-07-23 1
182 1977-07-31 1
183 1977-10-04 1
184 1978-01-04 1
185 1978-04-06 1
186 1978-05-27 1
187 1978-06-08 1
188 1978-08-04 1
189 1978-10-03 1
190 1978-12-15 1
191 1978-12-17 1
192 1979-01-31 1
193 1979-05-07 1
194 1979-06-16 1
195 1979-06-24 1
196 1979-07-14 1
197 1979-08-29 1
198 1979-09-17 1
199 1979-09-20 1
200 1979-09-22 1
201 1979-10-28 1
202 1979-12-25 1
203 1979-12-29 1
204 1980-01-05 1
205 1980-02-11 1
206 1980-02-28 1
207 1980-03-08 1
208 1980-03-14 1
209 1980-04-01 1
210 1980-08-10 1
211 1980-09-23 1
212 1980-11-12 1
213 1980-11-16 1
214 1980-11-19 1
215 1980-11-29 1
216 1980-12-24 1
217 1981-01-24 1
218 1981-02-17 1
219 1981-04-09 1
220 1981-05-22 1
221 1981-06-09 1
222 1981-08-03 1

223 1981-08-05 1
224 1981-09-26 1
225 1981-10-08 1
226 1981-10-26 1
227 1981-10-29 1
228 1982-03-02 1
229 1982-04-11 1
230 1982-06-08 1
231 1982-08-19 1
232 1982-08-31 1
233 1982-09-10 1
234 1982-10-19 1
235 1982-12-16 1
236 1983-03-18 1
237 1983-04-11 1
238 1983-07-29 1
239 1983-10-19 1
240 1983-11-01 1
241 1983-11-02 1
242 1984-01-27 1
243 1984-03-22 1
244 1984-05-28 1
245 1984-07-16 1
246 1984-07-24 1
247 1984-10-05 1
248 1984-10-29 1
249 1984-11-13 1
250 1984-12-11 1
251 1984-12-28 1
252 1985-01-10 1
253 1985-01-17 1
254 1985-02-12 1
255 1985-02-27 1
256 1985-03-05 1
257 1985-03-22 1
258 1985-09-12 1
259 1985-09-16 1
260 1985-12-19 1
261 1985-12-28 1
262 1986-01-02 1
263 1986-02-23 1
264 1986-03-21 1
265 1986-03-31 1
266 1986-04-22 1
267 1986-08-13 1
268 1986-11-14 1
269 1986-11-15 1
270 1986-12-25 1
271 1987-03-04 1
272 1987-05-13 1
273 1987-05-18 1
274 1987-07-03 1
275 1987-09-19 1
276 1987-10-23 1
277 1987-10-26 1
278 1987-11-23 1

279 1987-11-24 1
280 1988-04-02 1
281 1988-04-13 1
282 1988-05-02 1
283 1988-05-19 1
284 1988-07-10 1
285 1988-08-17 1
286 1988-09-12 1
287 1988-09-17 1
288 1988-10-05 1
289 1988-11-13 1
290 1988-12-08 1
291 1989-02-15 1
292 1989-03-16 1
293 1989-07-14 1
294 1990-01-06 1
295 1990-01-18 1
296 1990-02-05 1
297 1990-02-06 1
298 1990-03-15 1
299 1990-04-17 1
300 1990-04-26 1
301 1990-05-21 1
302 1990-07-05 1
303 1990-08-07 1
304 1990-09-22 1
305 1990-10-26 1
306 1990-12-12 1
307 1990-12-13 1
308 1991-01-05 1
309 1991-02-24 1
310 1991-04-09 1
311 1991-04-21 1
312 1991-05-01 1
313 1991-06-22 1
314 1991-08-03 1
315 1991-09-23 1
316 1991-10-01 1
317 1991-10-04 1
318 1991-11-27 1
319 2022-03-24 1
320 2022-03-30 1
321 2022-04-13 1
322 2022-04-15 1
323 2022-06-18 1
324 2022-08-12 1
325 2022-10-09 1
326 2022-10-13 1
327 2022-12-06 1
328 2022-12-15 1
329 2023-01-03 1
330 2023-01-13 1
331 2023-01-18 1
332 2023-01-30 1
333 2023-02-01 1
334 2023-02-14 1

335 2023-02-19 1
336 2023-03-12 1
337 2023-03-16 1
338 2023-04-11 1
339 2023-05-21 1
340 2023-06-21 1
341 2023-07-24 1
342 2023-08-08 1
343 2023-09-07 1
344 2023-09-25 1
345 2023-09-26 1
346 2023-10-03 1
347 2023-10-06 1
348 2023-11-10 1
349 2023-12-07 1
350 2024-01-08 1
351 2024-01-13 1
352 2024-01-23 1
353 2024-02-23 1
354 2024-03-15 1
355 2024-04-13 1
356 2024-06-11 1
357 2024-06-21 1
358 2024-06-30 1
359 2024-07-19 1
360 2024-07-23 1
361 2024-08-31 1
362 2024-09-17 1
363 2024-10-07 1
364 2024-10-25 1
365 2024-11-19 1
366 2024-12-08 1
367 2024-12-13 1
368 2024-12-18 1
369 2025-01-28 1
370 2025-03-11 1
371 2025-05-04 1
372 2025-08-12 1
373 2026-05-28 1
374 2026-07-11 1
375 2026-07-19 1
376 2026-08-05 1
377 2026-08-24 1
378 2026-08-27 1
379 2026-09-17 1
380 2026-09-27 1
381 2026-10-18 1
382 2026-11-10 1
383 2026-11-18 1
384 2026-11-20 1
385 2026-11-23 1
386 2027-02-04 1
387 2027-03-03 1
388 2027-03-05 1
389 2027-04-06 1
390 2027-04-16 1

391 2027-04-20 1
392 2027-04-27 1
393 2027-05-14 1
394 2027-07-10 1
395 2027-09-12 1
396 2027-11-04 1
397 2027-11-06 1
398 2027-12-28 1
399 2028-01-04 1
400 2028-01-12 1
401 2028-01-19 1
402 2028-02-25 1
403 2028-03-03 1
404 2028-03-18 1
405 2028-03-20 1
406 2028-03-22 1
407 2028-03-28 1
408 2028-04-09 1
409 2028-05-14 1
410 2028-07-03 1
411 2028-07-25 1
412 2028-08-25 1
413 2028-08-26 1
414 2028-09-24 1
415 2028-11-23 1
416 2029-01-06 1
417 2029-01-19 1
418 2029-02-25 1
419 2029-04-01 1
420 2029-06-02 1
421 2029-06-10 1
422 2029-07-05 1
423 2029-07-07 1
424 2029-08-23 1
425 2029-08-28 1
426 2029-09-18 1
427 2029-09-30 1
428 2029-10-25 1
429 2029-10-29 1
430 2029-12-30 1
431 2030-01-15 1
432 2030-02-09 1
433 2030-03-06 1
434 2030-05-20 1
435 2030-05-24 1
436 2030-07-24 1
437 2030-08-17 1
438 2030-09-07 1
439 2030-09-20 1
440 2030-10-12 1
441 2030-10-14 1
442 2030-10-31 1
443 2031-01-01 1
444 2031-01-12 1
445 2031-01-28 1
446 2031-02-10 1

447 2031-02-12 1
448 2031-03-27 1
449 2031-04-10 1
450 2031-04-21 1
451 2031-06-01 1
452 2031-06-22 1
453 2031-07-05 1
454 2031-08-26 1
455 2031-09-05 1
456 2031-10-21 1
457 2031-12-04 1
458 2031-12-13 1
459 2031-12-15 1
460 2031-12-18 1
461 2031-12-28 1
462 2032-01-16 1
463 2032-03-06 1
464 2032-03-08 1
465 2032-03-23 1
466 2032-03-31 1
467 2032-05-25 1
468 2032-06-14 1
469 2032-06-17 1
470 2032-07-11 1
471 2032-07-15 1
472 2032-09-03 1
473 2032-09-07 1
474 2032-09-17 1
475 2032-10-14 1
476 2032-10-24 1
477 2032-10-27 1
478 2032-11-22 1
479 2033-01-07 1
480 2033-01-31 1
481 2033-03-27 1
482 2033-05-08 1
483 2033-05-10 1
484 2033-05-15 1
485 2033-07-01 1
486 2033-07-14 1
487 2033-09-12 1
488 2033-10-28 1
489 2033-11-04 1
490 2033-11-25 1
491 2033-12-09 1
492 2033-12-16 1
493 2034-01-05 1
494 2034-03-27 1
495 2034-06-06 1
496 2034-06-09 1
497 2034-07-15 1
498 2034-10-07 1
499 2034-12-24 1
500 2035-01-16 1
501 2035-01-20 1
502 2035-01-21 1

503 2035-02-03 1
504 2035-02-20 1
505 2035-03-04 1
506 2035-03-11 1
507 2035-03-29 1
508 2035-04-13 1
509 2035-05-18 1
510 2035-06-14 1
511 2035-06-17 1
512 2035-07-05 1
513 2035-08-09 1
514 2035-09-12 1
515 2035-09-25 1
516 2035-10-01 1
517 2035-10-10 1
518 2035-10-16 1
519 2035-11-03 1
520 2035-11-10 1
521 2036-02-18 1
522 2036-04-01 1
523 2036-05-25 1
524 2036-07-21 1
525 2036-10-27 1
526 2036-11-01 1
527 2037-01-05 1
528 2037-03-02 1
529 2037-03-05 1
530 2037-03-27 1
531 2037-08-25 1
532 2037-10-09 1
533 2037-12-25 1
534 2037-12-30 1
535 2038-01-02 1
536 2038-03-26 1
537 2038-04-24 1
538 2038-07-05 1
539 2038-07-14 1
540 2038-07-21 1
541 2038-08-31 1
542 2038-10-05 1
543 2038-10-17 1
544 2038-11-24 1
545 2039-03-15 1
546 2039-04-02 1
547 2039-05-07 1
548 2039-05-16 1
549 2039-06-02 1
550 2039-09-08 1
551 2039-09-21 1
552 2039-09-28 1
553 2039-10-18 1
554 2039-10-24 1
555 2039-10-30 1
556 2039-11-04 1
557 2040-01-01 1
558 2040-02-14 1

559 2040-02-19 1
560 2040-02-20 1
561 2040-03-28 1
562 2040-04-01 1
563 2040-04-10 1
564 2040-05-13 1
565 2040-06-04 1
566 2040-07-06 1
567 2040-07-23 1
568 2040-09-11 1
569 2040-10-13 1
570 2040-10-16 1
571 2040-10-27 1
572 2040-10-31 1
573 2041-02-02 1
574 2041-02-18 1
575 2041-02-20 1
576 2041-02-27 1
577 2041-03-18 1
578 2041-04-01 1
579 2041-04-07 1
580 2041-04-25 1
581 2041-04-28 1
582 2041-05-09 1
583 2041-05-10 1
584 2041-06-19 1
585 2041-08-04 1
586 2041-08-31 1
587 2041-09-27 1
588 2041-11-10 1
589 2041-11-28 1
590 2041-12-08 1
591 2041-12-20 1
592 2042-04-06 1
593 2042-06-01 1
594 2042-06-11 1
595 2042-06-16 1
596 2042-07-27 1
597 2042-08-27 1
598 2042-11-09 1
599 2043-04-13 1
600 2043-04-18 1
601 2043-05-27 1
602 2043-07-03 1
603 2043-07-28 1
604 2043-09-12 1
605 2043-09-14 1
606 2043-09-30 1
607 2043-10-04 1
608 2043-11-24 1
609 2043-12-12 1
610 2044-01-21 1
611 2044-02-21 1
612 2044-03-08 1
613 2044-04-14 1
614 2044-06-22 1

615 2044-07-22 1
616 2044-08-10 1
617 2044-09-12 1
618 2044-10-01 1
619 2044-10-16 1
620 2044-11-14 1
621 2045-03-04 1
622 2045-03-24 1
623 2045-04-25 1
624 2045-04-28 1
625 2045-06-22 1
626 2045-07-31 1
627 2045-08-26 1
628 2045-09-03 1
629 2045-09-10 1
630 2045-10-15 1
631 2045-11-12 1
632 2045-12-06 1
633 2045-12-13 1
634 2045-12-27 1
635 2046-01-13 1
636 2046-01-16 1
637 2046-02-01 1
638 2046-03-06 1
639 2046-03-14 1
640 2046-04-02 1
641 2046-04-25 1
642 2046-05-16 1
643 2046-06-06 1
644 2046-06-22 1
645 2046-08-21 1
646 2046-09-12 1
647 2047-01-02 1
648 2047-02-04 1
649 2047-02-07 1
650 2047-03-30 1
651 2047-04-25 1
652 2047-05-13 1
653 2047-05-22 1
654 2047-07-07 1
655 2047-08-09 1
656 2047-08-18 1
657 2047-10-17 1
658 2047-11-03 1
659 2047-11-11 1
660 2047-11-25 1
661 2047-12-10 1
662 2048-02-28 1
663 2048-03-28 1
664 2048-04-05 1
665 2048-05-30 1
666 2048-08-21 1
667 2048-11-10 1
668 2048-12-17 1
669 2049-02-16 1
670 2049-02-27 1

671 2049-04-10 1
672 2049-04-23 1
673 2049-09-12 1
674 2050-02-16 1
675 2050-04-05 1
676 2050-04-14 1
677 2050-04-29 1
678 2050-06-15 1
679 2050-07-28 1
680 2050-08-19 1
681 2050-11-19 1
682 2050-12-26 1
683 2051-03-24 1
684 2051-03-31 1
685 2051-04-04 1
686 2051-04-12 1
687 2051-06-04 1
688 2051-06-11 1
689 2051-06-24 1
690 2051-09-18 1
691 2051-09-24 1
692 2051-11-22 1
693 2052-01-15 1
694 2052-01-23 1
695 2052-02-22 1
696 2052-02-25 1
697 2052-04-05 1
698 2052-05-17 1
699 2052-06-28 1
700 2052-06-30 1
701 2052-08-31 1
702 2052-09-11 1
703 2052-10-31 1
704 2052-11-10 1
705 2052-11-12 1
706 2052-12-06 1
707 2052-12-08 1
708 2053-03-14 1
709 2053-03-18 1
710 2053-04-04 1
711 2053-04-17 1
712 2053-04-21 1
713 2053-05-03 1
714 2053-06-06 1
715 2053-10-13 1
716 2053-10-25 1
717 2053-11-09 1
718 2053-11-12 1
719 2053-11-23 1
720 2053-12-04 1
721 2053-12-25 1
722 2053-12-27 1
723 2054-02-03 1
724 2054-03-06 1
725 2054-03-27 1
726 2054-04-14 1

727 2054-05-01 1
728 2054-07-17 1
729 2054-07-23 1
730 2054-09-25 1
731 2054-10-04 1
732 2054-10-16 1
733 2055-01-05 1
734 2055-01-26 1
735 2055-04-10 1
736 2055-04-13 1
737 2055-09-04 1
738 2055-10-09 1
739 2055-10-11 1
740 2055-12-04 1
741 2055-12-13 1
742 2056-01-19 1
743 2056-01-30 1
744 2056-02-12 1
745 2056-03-13 1
746 2056-03-18 1
747 2056-07-05 1
748 2056-07-09 1
749 2056-07-18 1
750 2056-10-26 1
751 2056-12-28 1
752 2056-12-29 1
753 2057-01-14 1
754 2057-03-17 1
755 2057-06-28 1
756 2057-07-21 1
757 2057-08-21 1
758 2057-10-14 1
759 2057-11-08 1
760 2057-11-15 1
761 2057-12-27 1
762 2058-02-16 1
763 2058-03-21 1
764 2058-03-25 1
765 2058-04-07 1
766 2058-05-08 1
767 2058-06-04 1
768 2058-07-13 1
769 2058-08-03 1
770 2058-08-29 1
771 2058-11-10 1
772 2058-11-12 1
773 2059-01-12 1
774 2059-02-14 1
775 2059-02-24 1
776 2059-03-04 1
777 2059-05-13 1
778 2059-06-08 1
779 2059-06-14 1
780 2059-07-04 1
781 2059-07-28 1
782 2059-10-10 1

783 2059-11-09 1
784 2059-11-15 1
785 2059-11-24 1
786 2059-11-28 1
787 2059-12-02 1
788 2060-01-18 1
789 2060-02-20 1
790 2060-03-17 1
791 2060-03-19 1
792 2060-03-25 1
793 2060-06-18 1
794 2060-07-26 1
795 2060-08-20 1
796 2060-10-11 1
797 2060-10-23 1
798 2060-11-16 1
799 2060-12-19 1
800 2061-01-04 1
801 2061-03-03 1
802 2061-03-28 1
803 2061-03-30 1
804 2061-06-05 1
805 2061-06-17 1
806 2061-07-07 1
807 2061-08-06 1
808 2061-10-28 1
809 2062-02-12 1
810 2062-02-17 1
811 2062-06-09 1
812 2062-08-08 1
813 2062-11-25 1
814 2062-12-05 1
815 2062-12-11 1
816 2063-01-20 1
817 2063-02-10 1
818 2063-03-09 1
819 2063-05-01 1
820 2063-05-08 1
821 2063-08-15 1
822 2063-08-27 1
823 2063-11-20 1
824 2064-03-15 1
825 2064-04-12 1
826 2064-04-22 1
827 2064-04-23 1
828 2064-04-24 1
829 2064-05-05 1
830 2064-06-01 1
831 2064-06-15 1
832 2064-07-19 1
833 2064-07-26 1
834 2064-07-28 1
835 2064-08-21 1
836 2064-09-16 1
837 2064-10-06 1
838 2064-10-15 1

```

839 2065-02-25 1
840 2065-04-22 1
841 2065-05-07 1
842 2065-05-15 1
843 2065-08-19 1
844 2065-09-12 1
845 2065-12-08 1
846 2066-01-05 1
847 2066-04-01 1
848 2066-04-04 1
849 2066-04-08 1
850 2066-04-18 1
851 2066-06-01 1
852 2066-08-18 1
853 2066-09-23 1
854 2066-09-25 1
855 2066-10-20 1
856 2066-11-11 1
857 2066-11-21 1
858 2066-11-29 1
859 2066-12-08 1
860 2066-12-30 1
861 2066-12-31 1
862 2067-02-02 1
863 2067-03-24 1
864 2067-04-03 1
865 2067-05-02 1
866 2067-05-08 1
867 2067-06-14 1
868 2067-07-21 1
869 2067-08-20 1
870 2067-09-06 1
871 2067-10-22 1
872 2067-11-21 1
873 2068-01-08 1
874 2068-01-27 1
875 2068-05-16 1
876 2068-05-21 1
877 2068-07-17 1
878 2068-08-05 1
879 2068-09-15 1
880 2068-10-08 1

```

```

### DEATHDATE
      DEATHDATE  n
1      <NA> 820
2  2012-01-26   2
3  2014-10-19   2
4  2016-01-31   2
5  2017-09-29   2
6  2019-07-31   2
7  2021-01-05   2
8  2011-02-03   1
9  2011-02-17   1
10 2011-07-12   1
11 2011-10-10   1

```

12	2011-11-29	1
13	2011-12-08	1
14	2012-01-22	1
15	2012-02-02	1
16	2012-03-08	1
17	2012-05-01	1
18	2012-05-15	1
19	2012-07-03	1
20	2012-07-11	1
21	2012-09-29	1
22	2012-10-17	1
23	2012-11-22	1
24	2012-12-09	1
25	2012-12-23	1
26	2012-12-29	1
27	2013-04-29	1
28	2013-05-14	1
29	2013-06-02	1
30	2013-06-10	1
31	2013-06-26	1
32	2013-07-04	1
33	2013-07-25	1
34	2013-08-29	1
35	2013-10-27	1
36	2013-11-12	1
37	2013-11-28	1
38	2013-12-11	1
39	2014-01-13	1
40	2014-01-14	1
41	2014-02-11	1
42	2014-02-13	1
43	2014-02-14	1
44	2014-02-23	1
45	2014-04-21	1
46	2014-07-19	1
47	2014-07-29	1
48	2014-11-10	1
49	2014-11-16	1
50	2014-11-27	1
51	2014-12-21	1
52	2015-02-02	1
53	2015-02-19	1
54	2015-03-31	1
55	2015-04-30	1
56	2015-05-05	1
57	2015-07-29	1
58	2016-01-02	1
59	2016-04-04	1
60	2016-04-09	1
61	2016-04-22	1
62	2016-07-17	1
63	2016-08-31	1
64	2016-10-03	1
65	2016-10-06	1
66	2016-10-21	1
67	2016-11-06	1

68	2016-12-17	1
69	2016-12-30	1
70	2017-01-19	1
71	2017-02-04	1
72	2017-02-13	1
73	2017-03-01	1
74	2017-04-03	1
75	2017-04-10	1
76	2017-04-12	1
77	2017-04-27	1
78	2017-07-16	1
79	2017-07-25	1
80	2017-08-07	1
81	2017-09-07	1
82	2017-09-19	1
83	2017-09-20	1
84	2017-12-16	1
85	2018-02-18	1
86	2018-03-06	1
87	2018-03-29	1
88	2018-04-06	1
89	2018-04-08	1
90	2018-05-01	1
91	2018-05-04	1
92	2018-05-05	1
93	2018-05-16	1
94	2018-05-24	1
95	2018-05-27	1
96	2018-06-05	1
97	2018-06-24	1
98	2018-07-14	1
99	2018-07-21	1
100	2018-08-17	1
101	2018-09-25	1
102	2018-10-05	1
103	2018-10-31	1
104	2018-12-17	1
105	2018-12-28	1
106	2019-02-06	1
107	2019-02-12	1
108	2019-03-17	1
109	2019-03-18	1
110	2019-03-22	1
111	2019-03-26	1
112	2019-05-21	1
113	2019-05-31	1
114	2019-08-22	1
115	2019-08-28	1
116	2019-09-23	1
117	2019-10-06	1
118	2019-10-28	1
119	2019-11-08	1
120	2019-11-25	1
121	2019-11-30	1
122	2020-01-31	1
123	2020-04-25	1

124	2020-05-06	1
125	2020-06-02	1
126	2020-07-09	1
127	2020-08-02	1
128	2020-08-29	1
129	2020-10-15	1
130	2020-10-27	1
131	2020-11-15	1
132	2020-11-28	1
133	2020-12-07	1
134	2020-12-09	1
135	2021-01-02	1
136	2021-01-04	1
137	2021-01-13	1
138	2021-01-19	1
139	2021-01-22	1
140	2021-01-27	1
141	2021-03-08	1
142	2021-03-11	1
143	2021-03-26	1
144	2021-05-01	1
145	2021-05-04	1
146	2021-05-15	1
147	2021-10-17	1
148	2021-12-26	1
149	2022-01-27	1

CITY

	CITY	n
1	Boston	541
2	Quincy	80
3	Cambridge	45
4	Revere	42
5	Chelsea	39
6	Weymouth	37
7	Somerville	25
8	Hingham	22
9	Winthrop	22
10	Brookline	17
11	Everett	16
12	Hull	15
13	Medford	13
14	Braintree	10
15	Cohasset	10
16	Malden	8
17	Scituate	8
18	Newton	6
19	Stoneham	5
20	North Scituate	3
21	Reading	2
22	Belmont	1
23	Lynnfield	1
24	Melrose	1
25	Milton	1
26	Norwell	1
27	Waltham	1

```
28      Watertown    1
29      Winchester    1
```

STATE

```
      STATE    n
1 Massachusetts 974
```

MARITAL

```
      MARITAL    n
1          M 782
2          S 189
3      Fine     1
4      male     1
5      <NA>     1
```

RACE

```
      RACE    n
1   white 680
2   black 163
3   asian  90
4   other  16
5 hawaiian 13
6   native 11
7   asiann  1
```

ETHNICITY

```
      ETHNICITY    n
1 nonhispanic 781
2   hispanic 190
3 nonhispani   2
4   hispani   1
```

GENDER

```
      GENDER    n
1          M 493
2          F 478
3 Female     1
4   Male     1
5 female     1
```

Question 10

Variables to clean:

MARITAL: 1 "Fine", 1 "male", 1 NA. Trim the illogical values.

RACE: 1 "asiann". Combine into "asian"

ETHNICITY: 1 "nonhispani", 1 "hispani". Combine into "nonhispanic" and "hispanic" respectively.

GENDER: 1 Female, 1 female, 1 Male. Combine into appropriate F or M category

```
In [9]: pat_join <- pat_join %>%
  filter(!(MARITAL %in% c("Fine", "male")))
pat_join$RACE[pat_join$RACE %in% c("asiann")] <- "asian"
pat_join$ETHNICITY[pat_join$ETHNICITY %in% c("nonhispani")] <- "nonhispanic"
```

```
pat_join$ETHNICITY[pat_join$ETHNICITY %in% c("hispani")] <- "hispanic"
pat_join$GENDER[pat_join$GENDER %in% c("female", "Female")] <- "F"
pat_join$GENDER[pat_join$GENDER %in% c("Male")] <- "M"
```

```
In [10]: pat_join <- pat_join %>%
  mutate(
    BIRTHDATE = if_else(year(BIRTHDATE) >= 2000 & year(BIRTHDATE) <= 2099,
      BIRTHDATE %m-% years(100),
      BIRTHDATE)
  )

head(pat_join, 10)
```


A data.frame: 10 × 11

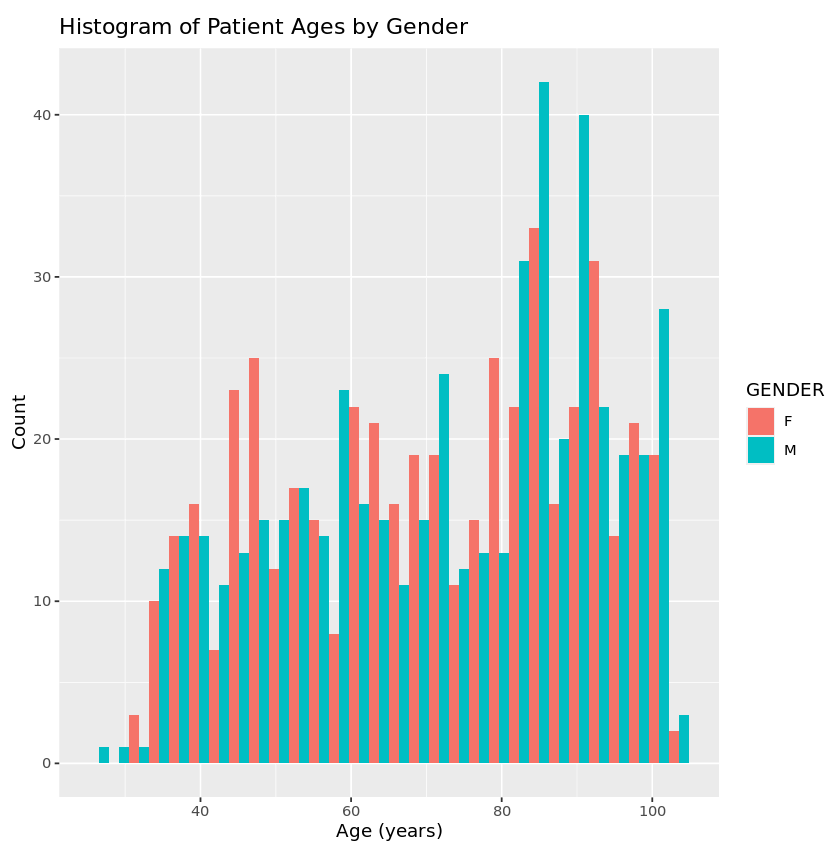
	ID	BIRTHDATE	DEATHDATE	FIRST	LAST	CITY	STATE
	<chr>	<date>	<date>	<chr>	<chr>	<chr>	<chr>
1	5605b66b-e92d-c16c-1b83-b8bf7040d51f	1977-03-19	NA	Nikita	Erdman	Quincy	Massachusetts
2	6e5ae27c-8038-7988-e2c0-25a103f01bfa	1940-02-19	NA	Zane	Hodkiewicz	Boston	Massachusetts
3	8123d076-0886-9007-e956-d5864aa121a7	1958-06-04	NA	Quinn	Marquardt	Quincy	Massachusetts
4	770518e4-6133-648e-60c9-071eb2f0e2ce	1928-12-25	2017-09-29	Abel	Smitham	Boston	Massachusetts
5	f96addf5-81b9-0aab-7855-d208d3d352c5	1928-12-25	2014-02-23	Edwin	Labadie	Boston	Massachusetts
6	8e9650d1-788a-78f9-4a28-d08f7f95354a	1928-12-25	NA	Frankie	Oberbrunner	Boston	Massachusetts
7	183df435-4190-060e-8f8e-bf63c572b266	1957-11-08	NA	Eilene	Walsh	Cambridge	Massachusetts
8	720560d4-51da-c38c-ee90-c15935278df1	1972-06-27	NA	Lowell	Price	Quincy	Massachusetts
9	217851b0-5f47-d376-18b9-0fe4ba77207e	1954-03-06	NA	Adrian	Gleason	Boston	Massachusetts
10	ff331e5c-ab16-e218-f39a-63e11de1ed75	1927-07-10	NA	Eugene	Abernathy	Boston	Massachusetts



Question 11

```
In [11]: pat_join <- pat_join %>%
  mutate(
    age = ifelse(
      is.na(DEATHDATE),
      floor(as.numeric(difftime(Sys.Date(), BIRTHDATE, units = "days"))) / 365
      floor(as.numeric(difftime(DEATHDATE, BIRTHDATE, units = "days"))) / 365.2
    )
  )
```

```
In [12]: ggplot(pat_join, aes(x = age, fill = GENDER)) +
  geom_histogram(position = "dodge", bins = 30) +
  labs(
    title = "Histogram of Patient Ages by Gender",
    x = "Age (years)",
    y = "Count"
  )
```



```
In [13]: ggplot(pat_join, aes(x = MARITAL, y = BIRTHDATE, color = MARITAL)) +
  geom_jitter(alpha = 0.7, width = 0.35, height = 0.2) +
  labs(
    title = "Birthdates by Marital Status",
    x = "Marital Status",
    y = "Birthdate"
  )
```

