

# Homework 03 - Nonstandard Evaluation and Git

## Nonstandard Evaluation

### Question 1

Imagine we have a data frame called `data`, with a `type` column. Which one works and why?

Function 1:

```
group_and_tally <- function(df, column){  
  df %>% group_by({{ column }}) %>% tally();  
}  
group_and_tally(data, type);
```

Function 2:

```
group_and_tally <- function(df, column){  
  df %>% group_by(column) %>% tally();  
}  
group_and_tally(data, type);
```

In [ ]: Function 1 would work because the curly curly parenthesis help to group data meanwhile the `(column)` will group the data by a column named `"column"`. Since the function will not be able to be performed.

## Git

For the questions below, please add the commands you used to complete these steps.

### Question 2

Set up your git repo on your local computer. If you already make a git repo on GitHub, but it isn't on your local computer - clone it.

In [ ]: `git clone git@github.com:nambial1/my_first_binder.git`

### Question 3

Set up your SSH key.

```
In [ ]: ssh-keygen -t ed25519 -C "nambia1" -f ~/.ssh/github_key

eval "$(ssh-agent -s)"
ssh-add ~/.ssh/github_key

cat ~/.ssh/github_key.pub
```

## Question 4

a) Add a HW2 directory to your git repo through the terminal with a HW.md file that says "This is for homework 2."

```
mkdir HW2 echo "This is for Homework2." >HW2/HW2.md
```

b) Add HW2.md to the staging area. Then, use the command to see which files have been modified, staged for commit, or are untracked. What does it show? They should copy paste the terminal response after git status, and show that key used the commands below.

```
In [ ]: git add HW2/HW2.md
git status
```

c) Save file changes to the main branch.

```
In [ ]: git commit -m "Add HW folder and HW2.md file"
git push
```

d) Now, edit the HW2.md file to give it a title.

```
In [ ]: echo "#Homework2" >> HW2/HW2.md
```

e) Use the command that compares current, unsaved changes to the main branch. What does it say?

```
git diff
```

f) Use the command that checks the status of the working directory and the staging area *again*. What does it say?

```
git status
```

g) Once again, add HW2.md to the staging area and save the file changes to the main branch. Then, get use the command that gives you project history and paste the output in your homework.

```
git add HW2.HW2.md git commit -m "Add title to HW2.md" git log
```

h) Do some searching... What `git` command will provide you documentation on other commands? Use that command to find documentation on `git log` and `git show`. What does `--since` mean in regards to `git log`? Copy and paste what is written in the documentation.

```
In [ ]: Changes to be committed:
        (use "git restore --staged <file>..." to unstage)
         new file:   HW2/patient_names.csv
         new file:   HW2/patient_properties.csv

Untracked files:
        (use "git add <file>..." to include in what will be committed)
        .DS_Store
        HW2/.DS_Store

aditinambiar@mac my_first_binder % git commit "HW 2 data sets"
error: pathspec 'HW 2 data sets' did not match any file(s) known to git
aditinambiar@mac my_first_binder %
```

## Tidyverse

Note: Please make sure Binder is set up correctly to run this section. You can follow the instructions here: <https://github.com/rjenki/BIOS512>.

**Please show your code for this section!** Before completing this section, please run the following.

```
In [1]: 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      1.1.4      ✓ readr      2.1.5
✓ forcats    1.0.0      ✓ stringr    1.5.1
✓ ggplot2    3.5.2      ✓ tibble     3.3.0
✓ lubridate  1.9.4      ✓ tidyr      1.3.1
✓ purrr      1.0.4

— 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 c
onflicts to become errors
```

## Question 5

Download the `patient_names.csv` and `patient_properties.csv` files from Canvas and read them into R. Manually set the date columns to be date variables. Print the first 10

observations of each.

```
In [7]: patient_names <- read_csv("HW2/patient_names.csv")
patient_properties <- read_csv("HW2/patient_properties.csv")
```

**Rows: 974 Columns: 7**

— **Column specification** —

**Delimiter: ","**

**chr (7):** ID, BIRTHDATE, DEATHDATE, FIRST, LAST, CITY, STATE

**i** Use `spec()` to retrieve the full column specification for this data.

**i** Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

**Rows: 3896 Columns: 3**

— **Column specification** —

**Delimiter: ","**

**chr (3):** ID, property, value

**i** Use `spec()` to retrieve the full column specification for this data.

**i** Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

```
In [8]: patient_names$BIRTHDATE <- as.Date(patient_names$BIRTHDATE, format="%m/%d/%Y")
patient_names$DEATHDATE <- as.Date(patient_names$DEATHDATE, format="%m/%d/%Y")
```

```
In [9]: head(patient_names,10)
```

A tibble: 10 × 7

ID	BIRTHDATE	DEATHDATE	FIRST	LAST	CITY	
<chr>	<date>	<date>	<chr>	<chr>	<chr>	
5605b66b-e92d-c16c-1b83-b8bf7040d51f	77-03-19	NA	Nikita578	Erdman779	Quincy	Massachusetts
6e5ae27c-8038-7988-e2c0-25a103f01bfa	40-02-19	NA	Zane918	Hodkiewicz467	Boston	Massachusetts
8123d076-0886-9007-e956-d5864aa121a7	58-06-04	NA	Quinn173	Marquardt819	Quincy	Massachusetts
770518e4-6133-648e-60c9-071eb2f0e2ce	28-12-25	17-09-29	Abel832	Smitham825	Boston	Massachusetts
f96addf5-81b9-0aab-7855-d208d3d352c5	28-12-25	14-02-23	Edwin773	Labadie908	Boston	Massachusetts
8e9650d1-788a-78f9-4a28-d08f7f95354a	28-12-25	NA	Frankie174	Oberbrunner298	Boston	Massachusetts
183df435-4190-060e-8f8e-bf63c572b266	57-11-08	NA	Eilene124	Walsh511	Cambridge	Massachusetts
720560d4-51da-c38c-ee90-c15935278df1	72-06-27	NA	Lowell343	Price929	Quincy	Massachusetts
217851b0-5f47-d376-18b9-0fe4ba77207e	54-03-06	NA	Adrian111	Gleason633	Boston	Massachusetts
ff331e5c-ab16-e218-f39a-63e11de1ed75	27-07-10	NA	Eugene421	Abernathy524	Boston	Massachusetts

## Question 6

In the data frame pulled from `patient_properties`, you'll notice that the data is long, not wide. Do a pivot to make the properties their own columns. Print the first 10 observations

after you do so.

```
In [10]: str(patient_properties)
         head(patient_properties, 10)
```

```
spec_tbl_ [3,896 × 3] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ ID      : chr [1:3896] "5605b66b-e92d-c16c-1b83-b8bf7040d51f" "5605b66b-e
92d-c16c-1b83-b8bf7040d51f" "5605b66b-e92d-c16c-1b83-b8bf7040d51f" "5605b66b
-e92d-c16c-1b83-b8bf7040d51f" ...
 $ property: chr [1:3896] "MARITAL" "RACE" "ETHNICITY" "GENDER" ...
 $ value   : chr [1:3896] "M" "white" "nonhispanic" "F" ...
- attr(*, "spec")=
 .. cols(
 ..   ID = col_character(),
 ..   property = col_character(),
 ..   value = col_character()
 .. )
- attr(*, "problems")=<externalptr>
      A tibble: 10 × 3
```

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

```
In [11]: library(tidyr)

patient_properties_wide <-patient_properties %>%
  pivot_wider(
    id_cols= ID,
    names_from= property,
    values_from= value
  )
```

```
In [12]: head (patient_properties_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

Perform a left join of the names and properties\_wide data frames by the ID column and print the first 10 rows.

```
In [13]: patients <- left_join(patient_names, patient_properties_wide, by = "ID")
          head(patients, 10)
```

A tibble: 10 × 11

	ID	BIRTHDATE	DEATHDATE	FIRST	LAST	CITY	
	<chr>	<date>	<date>	<chr>	<chr>	<chr>	
	5605b66b-e92d-c16c-1b83-b8bf7040d51f	77-03-19	NA	Nikita578	Erdman779	Quincy	Massac
	6e5ae27c-8038-7988-e2c0-25a103f01bfa	40-02-19	NA	Zane918	Hodkiewicz467	Boston	Massac
	8123d076-0886-9007-e956-d5864aa121a7	58-06-04	NA	Quinn173	Marquardt819	Quincy	Massac
	770518e4-6133-648e-60c9-071eb2f0e2ce	28-12-25	17-09-29	Abel832	Smitham825	Boston	Massac
	f96addf5-81b9-0aab-7855-d208d3d352c5	28-12-25	14-02-23	Edwin773	Labadie908	Boston	Massac
	8e9650d1-788a-78f9-4a28-d08f7f95354a	28-12-25	NA	Frankie174	Oberbrunner298	Boston	Massac
	183df435-4190-060e-8f8e-bf63c572b266	57-11-08	NA	Eilene124	Walsh511	Cambridge	Massac
	720560d4-51da-c38c-ee90-c15935278df1	72-06-27	NA	Lowell343	Price929	Quincy	Massac
	217851b0-5f47-d376-18b9-0fe4ba77207e	54-03-06	NA	Adrian111	Gleason633	Boston	Massac
	ff331e5c-ab16-e218-f39a-63e11de1ed75	27-07-10	NA	Eugene421	Abernathy524	Boston	Massac

## Question 8

Notice something interesting about the names in our data set. Fix the name formatting and print the first 10 observations.



```
In [14]: library (dplyr)
library (stringr)

patients <- patients %>%
  mutate(
    FIRST =str_trim(FIRST),
    LAST =str_trim(LAST),

    FIRST =str_remove(FIRST, "\\d+$"),
    LAST =str_remove(LAST, "\\d+$"),

    FIRST =str_to_title(FIRST),
    LAST =str_to_title(LAST)
  )
```

```
In [15]: head (patients, 10)
```

A tibble: 10 × 11

	ID	BIRTHDATE	DEATHDATE	FIRST	LAST	CITY	STATE
	<chr>	<date>	<date>	<chr>	<chr>	<chr>	<chr>
	5605b66b-e92d-c16c-1b83-b8bf7040d51f	77-03-19	NA	Nikita	Erdman	Quincy	Massachusetts
	6e5ae27c-8038-7988-e2c0-25a103f01bfa	40-02-19	NA	Zane	Hodkiewicz	Boston	Massachusetts
	8123d076-0886-9007-e956-d5864aa121a7	58-06-04	NA	Quinn	Marquardt	Quincy	Massachusetts
	770518e4-6133-648e-60c9-071eb2f0e2ce	28-12-25	17-09-29	Abel	Smitham	Boston	Massachusetts
	f96addf5-81b9-0aab-7855-d208d3d352c5	28-12-25	14-02-23	Edwin	Labadie	Boston	Massachusetts
	8e9650d1-788a-78f9-4a28-d08f7f95354a	28-12-25	NA	Frankie	Oberbrunner	Boston	Massachusetts
	183df435-4190-060e-8f8e-bf63c572b266	57-11-08	NA	Eilene	Walsh	Cambridge	Massachusetts
	720560d4-51da-c38c-ee90-c15935278df1	72-06-27	NA	Lowell	Price	Quincy	Massachusetts
	217851b0-5f47-d376-18b9-0fe4ba77207e	54-03-06	NA	Adrian	Gleason	Boston	Massachusetts
	ff331e5c-ab16-e218-f39a-63e11de1ed75	27-07-10	NA	Eugene	Abernathy	Boston	Massachusetts

## Question 9

Using a for statement to loop through the categorical variables (excluding name and ID), print the counts of each unique value in descending order, using the `mdpre()` function for

formatting.

```
In [16]: cat_vars <- setdiff(names(patients), c("ID", "FIRST", "LAST"))  
  
for (var in cat_vars){  
  mdpre(sort(table(patients[[var]]), decreasing = TRUE))  
}
```

25-11-17	27-04-22	29-01-28	29-05-15	36-01-12	25-08-10	25-11-16	26-08-07
4	4	4	4	4	3	3	3
27-02-05	28-11-04	28-12-25	30-11-03	33-08-24	22-05-31	23-09-22	25-01-15
3	3	3	3	3	2	2	2
25-02-09	25-05-10	25-06-01	25-11-20	26-07-14	26-12-27	27-06-04	27-09-08
2	2	2	2	2	2	2	2
27-10-16	29-06-30	29-10-14	30-04-25	30-12-14	31-02-27	31-05-04	32-09-12
2	2	2	2	2	2	2	2
32-11-06	33-05-23	33-08-13	33-09-02	34-10-28	35-02-15	35-04-16	35-06-25
2	2	2	2	2	2	2	2
36-02-07	36-06-11	36-06-26	36-11-08	37-09-04	38-02-26	39-07-05	39-09-23
2	2	2	2	2	2	2	2
41-12-18	42-05-18	42-06-18	42-08-03	46-01-09	46-08-05	48-03-05	50-12-23
2	2	2	2	2	2	2	2
51-03-01	51-05-03	52-04-01	52-11-02	54-01-14	54-08-26	57-01-19	58-06-29
2	2	2	2	2	2	2	2
60-09-13	63-09-24	64-01-05	64-07-01	66-11-09	72-04-25	73-10-28	73-11-17
2	2	2	2	2	2	2	2
76-12-11	77-01-10	82-08-09	84-06-14	22-03-24	22-03-30	22-04-13	22-04-15
2	2	2	2	1	1	1	1
22-06-18	22-08-12	22-10-09	22-10-13	22-12-06	22-12-15	23-01-03	23-01-13
1	1	1	1	1	1	1	1
23-01-18	23-01-30	23-02-01	23-02-14	23-02-19	23-03-12	23-03-16	23-04-11
1	1	1	1	1	1	1	1
23-05-21	23-06-21	23-07-24	23-08-08	23-09-07	23-09-25	23-09-26	23-10-03
1	1	1	1	1	1	1	1
23-10-06	23-11-10	23-12-07	24-01-08	24-01-13	24-01-23	24-02-23	24-03-15
1	1	1	1	1	1	1	1
24-04-13	24-06-11	24-06-21	24-06-30	24-07-19	24-07-23	24-08-31	24-09-17
1	1	1	1	1	1	1	1
24-10-07	24-10-25	24-11-19	24-12-08	24-12-13	24-12-18	25-01-28	25-03-11
1	1	1	1	1	1	1	1
25-05-04	25-08-12	26-05-28	26-07-11	26-07-19	26-08-05	26-08-24	26-08-27
1	1	1	1	1	1	1	1
26-09-17	26-09-27	26-10-18	26-11-10	26-11-18	26-11-20	26-11-23	27-02-04
1	1	1	1	1	1	1	1
27-03-03	27-03-05	27-04-06	27-04-16	27-04-20	27-04-27	27-05-14	27-07-10
1	1	1	1	1	1	1	1
27-09-12	27-11-04	27-11-06	27-12-28	28-01-04	28-01-12	28-01-19	28-02-25
1	1	1	1	1	1	1	1
28-03-03	28-03-18	28-03-20	28-03-22	28-03-28	28-04-09	28-05-14	28-07-03
1	1	1	1	1	1	1	1
28-07-25	28-08-25	28-08-26	28-09-24	28-11-23	29-01-06	29-01-19	29-02-25
1	1	1	1	1	1	1	1
29-04-01	29-06-02	29-06-10	29-07-05	29-07-07	29-08-23	29-08-28	29-09-18
1	1	1	1	1	1	1	1
29-09-30	29-10-25	29-10-29	29-12-30	30-01-15	30-02-09	30-03-06	30-05-20
1	1	1	1	1	1	1	1
30-05-24	30-07-24	30-08-17	30-09-07	30-09-20	30-10-12	30-10-14	30-10-31
1	1	1	1	1	1	1	1
31-01-01	31-01-12	31-01-28	31-02-10	31-02-12	31-03-27	31-04-10	31-04-21
1	1	1	1	1	1	1	1
31-06-01	31-06-22	31-07-05	31-08-26	31-09-05	31-10-21	31-12-04	31-12-13
1	1	1	1	1	1	1	1
31-12-15	31-12-18	31-12-28	32-01-16	32-03-06	32-03-08	32-03-23	32-03-31
1	1	1	1	1	1	1	1

32-05-25	32-06-14	32-06-17	32-07-11	32-07-15	32-09-03	32-09-07	32-09-17
1	1	1	1	1	1	1	1
32-10-14	32-10-24	32-10-27	32-11-22	33-01-07	33-01-31	33-03-27	33-05-08
1	1	1	1	1	1	1	1
33-05-10	33-05-15	33-07-01	33-07-14	33-09-12	33-10-28	33-11-04	33-11-25
1	1	1	1	1	1	1	1
33-12-09	33-12-16	34-01-05	34-03-27	34-06-06	34-06-09	34-07-15	34-10-07
1	1	1	1	1	1	1	1
34-12-24	35-01-16	35-01-20	35-01-21	35-02-03	35-02-20	35-03-04	35-03-11
1	1	1	1	1	1	1	1
35-03-29	35-04-13	35-05-18	35-06-14	35-06-17	35-07-05	35-08-09	35-09-12
1	1	1	1	1	1	1	1
35-09-25	35-10-01	35-10-10	35-10-16	35-11-03	35-11-10	36-02-18	36-04-01
1	1	1	1	1	1	1	1
36-05-25	36-07-21	36-10-27	36-11-01	37-01-05	37-03-02	37-03-05	37-03-27
1	1	1	1	1	1	1	1
37-08-25	37-10-09	37-12-25	37-12-30	38-01-02	38-03-26	38-04-24	38-07-05
1	1	1	1	1	1	1	1
38-07-14	38-07-21	38-08-31	38-10-05	38-10-17	38-11-24	39-03-15	39-04-02
1	1	1	1	1	1	1	1
39-05-07	39-05-16	39-06-02	39-09-08	39-09-21	39-09-28	39-10-18	39-10-24
1	1	1	1	1	1	1	1
39-10-30	39-11-04	40-01-01	40-02-14	40-02-19	40-02-20	40-03-28	40-04-01
1	1	1	1	1	1	1	1
40-04-10	40-05-13	40-06-04	40-07-06	40-07-23	40-09-11	40-10-13	40-10-16
1	1	1	1	1	1	1	1
40-10-27	40-10-31	41-02-02	41-02-18	41-02-20	41-02-27	41-03-18	41-04-01
1	1	1	1	1	1	1	1
41-04-07	41-04-25	41-04-28	41-05-09	41-05-10	41-06-19	41-08-04	41-08-31
1	1	1	1	1	1	1	1
41-09-27	41-11-10	41-11-28	41-12-08	41-12-20	42-04-06	42-06-01	42-06-11
1	1	1	1	1	1	1	1
42-06-16	42-07-27	42-08-27	42-11-09	43-04-13	43-04-18	43-05-27	43-07-03
1	1	1	1	1	1	1	1
43-07-28	43-09-12	43-09-14	43-09-30	43-10-04	43-11-24	43-12-12	44-01-21
1	1	1	1	1	1	1	1
44-02-21	44-03-08	44-04-14	44-06-22	44-07-22	44-08-10	44-09-12	44-10-01
1	1	1	1	1	1	1	1
44-10-16	44-11-14	45-03-04	45-03-24	45-04-25	45-04-28	45-06-22	45-07-31
1	1	1	1	1	1	1	1
45-08-26	45-09-03	45-09-10	45-10-15	45-11-12	45-12-06	45-12-13	45-12-27
1	1	1	1	1	1	1	1
46-01-13	46-01-16	46-02-01	46-03-06	46-03-14	46-04-02	46-04-25	46-05-16
1	1	1	1	1	1	1	1
46-06-06	46-06-22	46-08-21	46-09-12	47-01-02	47-02-04	47-02-07	47-03-30
1	1	1	1	1	1	1	1
47-04-25	47-05-13	47-05-22	47-07-07	47-08-09	47-08-18	47-10-17	47-11-03
1	1	1	1	1	1	1	1
47-11-11	47-11-25	47-12-10	48-02-28	48-03-28	48-04-05	48-05-30	48-08-21
1	1	1	1	1	1	1	1
48-11-10	48-12-17	49-02-16	49-02-27	49-04-10	49-04-23	49-09-12	50-02-16
1	1	1	1	1	1	1	1
50-04-05	50-04-14	50-04-29	50-06-15	50-07-28	50-08-19	50-11-19	50-12-26
1	1	1	1	1	1	1	1
51-03-24	51-03-31	51-04-04	51-04-12	51-06-04	51-06-11	51-06-24	51-09-18
1	1	1	1	1	1	1	1

51-09-24	51-11-22	52-01-15	52-01-23	52-02-22	52-02-25	52-04-05	52-05-17
1	1	1	1	1	1	1	1
52-06-28	52-06-30	52-08-31	52-09-11	52-10-31	52-11-10	52-11-12	52-12-06
1	1	1	1	1	1	1	1
52-12-08	53-03-14	53-03-18	53-04-04	53-04-17	53-04-21	53-05-03	53-06-06
1	1	1	1	1	1	1	1
53-10-13	53-10-25	53-11-09	53-11-12	53-11-23	53-12-04	53-12-25	53-12-27
1	1	1	1	1	1	1	1
54-02-03	54-03-06	54-03-27	54-04-14	54-05-01	54-07-17	54-07-23	54-09-25
1	1	1	1	1	1	1	1
54-10-04	54-10-16	55-01-05	55-01-26	55-04-10	55-04-13	55-09-04	55-10-09
1	1	1	1	1	1	1	1
55-10-11	55-12-04	55-12-13	56-01-19	56-01-30	56-02-12	56-03-13	56-03-18
1	1	1	1	1	1	1	1
56-07-05	56-07-09	56-07-18	56-10-26	56-12-28	56-12-29	57-01-14	57-03-17
1	1	1	1	1	1	1	1
57-06-28	57-07-21	57-08-21	57-10-14	57-11-08	57-11-15	57-12-27	58-02-16
1	1	1	1	1	1	1	1
58-03-21	58-03-25	58-04-07	58-05-08	58-06-04	58-07-13	58-08-03	58-08-29
1	1	1	1	1	1	1	1
58-11-10	58-11-12	59-01-12	59-02-14	59-02-24	59-03-04	59-05-13	59-06-08
1	1	1	1	1	1	1	1
59-06-14	59-07-04	59-07-28	59-10-10	59-11-09	59-11-15	59-11-24	59-11-28
1	1	1	1	1	1	1	1
59-12-02	60-01-18	60-02-20	60-03-17	60-03-19	60-03-25	60-06-18	60-07-26
1	1	1	1	1	1	1	1
60-08-20	60-10-11	60-10-23	60-11-16	60-12-19	61-01-04	61-03-03	61-03-28
1	1	1	1	1	1	1	1
61-03-30	61-06-05	61-06-17	61-07-07	61-08-06	61-10-28	62-02-12	62-02-17
1	1	1	1	1	1	1	1
62-06-09	62-08-08	62-11-25	62-12-05	62-12-11	63-01-20	63-02-10	63-03-09
1	1	1	1	1	1	1	1
63-05-01	63-05-08	63-08-15	63-08-27	63-11-20	64-03-15	64-04-12	64-04-22
1	1	1	1	1	1	1	1
64-04-23	64-04-24	64-05-05	64-06-01	64-06-15	64-07-19	64-07-26	64-07-28
1	1	1	1	1	1	1	1
64-08-21	64-09-16	64-10-06	64-10-15	65-02-25	65-04-22	65-05-07	65-05-15
1	1	1	1	1	1	1	1
65-08-19	65-09-12	65-12-08	66-01-05	66-04-01	66-04-04	66-04-08	66-04-18
1	1	1	1	1	1	1	1
66-06-01	66-08-18	66-09-23	66-09-25	66-10-20	66-11-11	66-11-21	66-11-29
1	1	1	1	1	1	1	1
66-12-08	66-12-30	66-12-31	67-02-02	67-03-24	67-04-03	67-05-02	67-05-08
1	1	1	1	1	1	1	1
67-06-14	67-07-21	67-08-20	67-09-06	67-10-22	67-11-21	68-01-08	68-01-27
1	1	1	1	1	1	1	1
68-05-16	68-05-21	68-07-17	68-08-05	68-09-15	68-10-08	69-01-11	69-01-26
1	1	1	1	1	1	1	1
69-02-10	69-02-14	69-02-17	69-03-21	69-05-06	69-05-15	69-07-12	69-11-20
1	1	1	1	1	1	1	1
69-12-27	69-12-30	70-04-08	70-04-16	70-06-03	70-09-08	70-09-17	70-11-26
1	1	1	1	1	1	1	1
70-12-02	71-01-26	71-01-30	71-03-31	71-05-10	71-05-16	71-06-19	71-07-05
1	1	1	1	1	1	1	1
71-07-08	71-08-19	71-08-26	71-10-27	71-11-06	72-01-15	72-01-25	72-04-15
1	1	1	1	1	1	1	1

72-06-10	72-06-27	72-06-29	72-07-30	72-10-24	73-01-06	73-01-24	73-02-16
1	1	1	1	1	1	1	1
73-02-26	73-03-12	73-04-19	73-05-16	73-05-25	73-06-12	73-08-28	73-09-01
1	1	1	1	1	1	1	1
73-09-09	74-01-20	74-01-30	74-04-21	74-04-22	74-04-27	74-05-12	74-06-19
1	1	1	1	1	1	1	1
74-07-09	74-07-15	74-07-29	74-10-03	74-12-06	74-12-22	75-01-24	75-02-05
1	1	1	1	1	1	1	1
75-02-22	75-03-13	75-04-29	75-05-07	75-05-20	75-05-31	75-06-27	75-07-27
1	1	1	1	1	1	1	1
75-09-12	75-09-19	75-11-14	75-12-06	75-12-30	76-01-07	76-01-22	76-02-27
1	1	1	1	1	1	1	1
76-04-04	76-04-10	76-05-18	76-05-26	76-08-01	76-08-18	76-09-02	76-10-13
1	1	1	1	1	1	1	1
76-10-18	76-10-31	76-12-01	77-01-02	77-01-30	77-02-05	77-02-08	77-03-19
1	1	1	1	1	1	1	1
77-04-05	77-04-19	77-04-29	77-05-16	77-05-28	77-06-29	77-07-23	77-07-31
1	1	1	1	1	1	1	1
77-10-04	78-01-04	78-04-06	78-05-27	78-06-08	78-08-04	78-10-03	78-12-15
1	1	1	1	1	1	1	1
78-12-17	79-01-31	79-05-07	79-06-16	79-06-24	79-07-14	79-08-29	79-09-17
1	1	1	1	1	1	1	1
79-09-20	79-09-22	79-10-28	79-12-25	79-12-29	80-01-05	80-02-11	80-02-28
1	1	1	1	1	1	1	1
80-03-08	80-03-14	80-04-01	80-08-10	80-09-23	80-11-12	80-11-16	80-11-19
1	1	1	1	1	1	1	1
80-11-29	80-12-24	81-01-24	81-02-17	81-04-09	81-05-22	81-06-09	81-08-03
1	1	1	1	1	1	1	1
81-08-05	81-09-26	81-10-08	81-10-26	81-10-29	82-03-02	82-04-11	82-06-08
1	1	1	1	1	1	1	1
82-08-19	82-08-31	82-09-10	82-10-19	82-12-16	83-03-18	83-04-11	83-07-29
1	1	1	1	1	1	1	1
83-10-19	83-11-01	83-11-02	84-01-27	84-03-22	84-05-28	84-07-16	84-07-24
1	1	1	1	1	1	1	1
84-10-05	84-10-29	84-11-13	84-12-11	84-12-28	85-01-10	85-01-17	85-02-12
1	1	1	1	1	1	1	1
85-02-27	85-03-05	85-03-22	85-09-12	85-09-16	85-12-19	85-12-28	86-01-02
1	1	1	1	1	1	1	1
86-02-23	86-03-21	86-03-31	86-04-22	86-08-13	86-11-14	86-11-15	86-12-25
1	1	1	1	1	1	1	1
87-03-04	87-05-13	87-05-18	87-07-03	87-09-19	87-10-23	87-10-26	87-11-23
1	1	1	1	1	1	1	1
87-11-24	88-04-02	88-04-13	88-05-02	88-05-19	88-07-10	88-08-17	88-09-12
1	1	1	1	1	1	1	1
88-09-17	88-10-05	88-11-13	88-12-08	89-02-15	89-03-16	89-07-14	90-01-06
1	1	1	1	1	1	1	1
90-01-18	90-02-05	90-02-06	90-03-15	90-04-17	90-04-26	90-05-21	90-07-05
1	1	1	1	1	1	1	1
90-08-07	90-09-22	90-10-26	90-12-12	90-12-13	91-01-05	91-02-24	91-04-09
1	1	1	1	1	1	1	1
91-04-21	91-05-01	91-06-22	91-08-03	91-09-23	91-10-01	91-10-04	91-11-27
1	1	1	1	1	1	1	1
12-01-26	14-10-19	16-01-31	17-09-29	19-07-31	21-01-05	11-02-03	11-02-17
2	2	2	2	2	2	1	1
11-07-12	11-10-10	11-11-29	11-12-08	12-01-22	12-02-02	12-03-08	12-05-01

1	1	1	1	1	1	1	1
12-05-15	12-07-03	12-07-11	12-09-29	12-10-17	12-11-22	12-12-09	12-12-23
1	1	1	1	1	1	1	1
12-12-29	13-04-29	13-05-14	13-06-02	13-06-10	13-06-26	13-07-04	13-07-25
1	1	1	1	1	1	1	1
13-08-29	13-10-27	13-11-12	13-11-28	13-12-11	14-01-13	14-01-14	14-02-11
1	1	1	1	1	1	1	1
14-02-13	14-02-14	14-02-23	14-04-21	14-07-19	14-07-29	14-11-10	14-11-16
1	1	1	1	1	1	1	1
14-11-27	14-12-21	15-02-02	15-02-19	15-03-31	15-04-30	15-05-05	15-07-29
1	1	1	1	1	1	1	1
16-01-02	16-04-04	16-04-09	16-04-22	16-07-17	16-08-31	16-10-03	16-10-06
1	1	1	1	1	1	1	1
16-10-21	16-11-06	16-12-17	16-12-30	17-01-19	17-02-04	17-02-13	17-03-01
1	1	1	1	1	1	1	1
17-04-03	17-04-10	17-04-12	17-04-27	17-07-16	17-07-25	17-08-07	17-09-07
1	1	1	1	1	1	1	1
17-09-19	17-09-20	17-12-16	18-02-18	18-03-06	18-03-29	18-04-06	18-04-08
1	1	1	1	1	1	1	1
18-05-01	18-05-04	18-05-05	18-05-16	18-05-24	18-05-27	18-06-05	18-06-24
1	1	1	1	1	1	1	1
18-07-14	18-07-21	18-08-17	18-09-25	18-10-05	18-10-31	18-12-17	18-12-28
1	1	1	1	1	1	1	1
19-02-06	19-02-12	19-03-17	19-03-18	19-03-22	19-03-26	19-05-21	19-05-31
1	1	1	1	1	1	1	1
19-08-22	19-08-28	19-09-23	19-10-06	19-10-28	19-11-08	19-11-25	19-11-30
1	1	1	1	1	1	1	1
20-01-31	20-04-25	20-05-06	20-06-02	20-07-09	20-08-02	20-08-29	20-10-15
1	1	1	1	1	1	1	1
20-10-27	20-11-15	20-11-28	20-12-07	20-12-09	21-01-02	21-01-04	21-01-13
1	1	1	1	1	1	1	1
21-01-19	21-01-22	21-01-27	21-03-08	21-03-11	21-03-26	21-05-01	21-05-04
1	1	1	1	1	1	1	1
21-05-15	21-10-17	21-12-26	22-01-27				
1	1	1	1				

Boston	Quincy	Cambridge	Revere	Chelsea
541	80	45	42	39
Weymouth	Somerville	Hingham	Winthrop	Brookline
37	25	22	22	17
Everett	Hull	Medford	Braintree	Cohasset
16	15	13	10	10
Malden	Scituate	Newton	Stoneham North	Scituate
8	8	6	5	3
Reading	Belmont	Lynnfield	Melrose	Milton
2	1	1	1	1
Norwell	Waltham	Watertown	Winchester	
1	1	1	1	
Massachusetts				
974				

M	S	Fine	male
782	189	1	1

white	black	asian	other	hawaiian	native	asiann
680	163	90	16	13	11	1



nonhispanic	hispanic	nonhispani	hispani
781	190	2	1

  

M	F female	Female	Male
493	478	1	1

## Question 10

If you see any weird values, get rid of the ones that don't make sense, and combine the ones that are formatted wrong. Don't forget to check the dates! Print the new tables for categorical values, and print the date ranges.

```
In [20]: patients <- patients %>%
  mutate(
    MARITAL =na_if(MARITAL, "Fine"),
    MARITAL =na_if(MARITAL, "male")
  )
```

```
In [24]: patients <- patients %>%
  mutate(
    GENDER = case_when(
      GENDER %in% c("Male", "male") ~ "M",
      GENDER %in% c("Female", "female") ~ "F",
      TRUE ~ GENDER)
  )
```

```
In [41]: patients <- patients %>%
  mutate(
    ETHNICITY = case_when(
      ETHNICITY %in% c("Nonhispani") ~ "Nonhispanic",
      ETHNICITY %in% c("Hispani") ~ "Hispanic",
      ETHNICITY %in% c("asiann") ~ "asian",
      TRUE ~ ETHNICITY)
  )
```

## Question 11

Make a histogram of the ages of patients by gender.

## Question 12

Make a scatterplot of birthdate by martial status.