

Intro to TCGA Clinical Data Homework #1 - Bruk Tefera

Bruk Tefera

07/17/22

Secure copy (scp) this file to your qbio_490_name repository. DO NOT EDIT IN sp25_course_materials!

(1) Set up your working environment

Before running this code, make sure you are in the correct working directory with `getwd()` and `setwd()`. Otherwise you'll have trouble tracking the data down!X

As good general practice to keep things organized, set the “analysis_data” folder as your working directory (“/PATH/TO/analysis_data”) whenever you are working with TCGA files. This is the same folder than contains the TCGA data you downloaded in the previous step. For most of you, it should be /home1/your_name/490_cluster/analysis_data.

Set your working directory below:

```
knitr::opts_knit$set(root.dir = normalizePath("/home1/tefera/490_cluster/analysis_data", mustWork = FALSE))
```

Before you move on, go to your GitHub web browser, open your .gitignore, and make sure the following files are listed.

```
#### TCGA Data
GDCdata
MANIFEST*
htseq_h5*
analysis_data"
```

The .gitignore file lets Github know which files to NOT push. The loaded files from TCGA are quite large, so we don't want to push them to GitHub. It's also typically best to avoid putting raw data on GitHub.

(1) Example data frames

Going back to variables and object types, data frames are objects that store two-dimensional representations of data, often with row and column names. Data frames are the bread and butter of data analysis in R!

Let's explore the built in mtcars data frame before we get to using the TCGA dataframe to learn about this crucial data type.

First, load in the data frame by running the following line of code. “mtcars” will then show up under “Data” in your environment panel.

```
data(mtcars)
```

The easiest way to view a data frame is to simply click on it in your environment tab! This will open up the matrix in a new tab. You can then explore the data visually.

However, this is not always helpful, especially when looking at larger or more complex data frames. We can use `head()` and `str()` to give us a better view in the console itself.

```
head(mtcars)
```

```
##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4      21.0   6  160 110 3.90 2.620 16.46  0  1   4    4
## Mazda RX4 Wag  21.0   6  160 110 3.90 2.875 17.02  0  1   4    4
## Datsun 710     22.8   4  108  93 3.85 2.320 18.61  1  1   4    1
## Hornet 4 Drive  21.4   6  258 110 3.08 3.215 19.44  1  0   3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0   3    2
## Valiant        18.1   6  225 105 2.76 3.460 20.22  1  0   3    1
```

Here we can see the first 6 rows of the mtcars data frame. We can see the row names (Mazda RX4, Mazda RX4 Wag, etc.), the column names (mpg, cyl, disp, etc.), and the data types associated with each column (dbl: double-precision floating point number).

```
str(mtcars) # str stands for structure
```

```
## 'data.frame':   32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Running str() will give us info about the structure of the data frame. We can see the data type (data.frame), the number of rows (obs.) and columns (variables), the general data type of each column (num), and a preview of the data in each column.

Note the “before each column name. In R, the dollar sign always refers to columns. We can view a singular column with the notation df_n `df_n$column_name`

Use this syntax to view the mpg column of the mtcars data frame.

```
mtcars$mpg
```

```
## [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2 10.4
## [16] 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8 19.7
## [31] 15.0 21.4
```

What if we want to view a column, but don't know it's exact name? We can get all column names (or all row names) by using the functions colnames() and rownames(). Call those below, then view the carburetor column using dollar sign notation.

```
colnames(mtcars)
```

```
## [1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am" "gear"
## [11] "carb"
```

```
rownames(mtcars)
```

```
## [1] "Mazda RX4"           "Mazda RX4 Wag"       "Datsun 710"
## [4] "Hornet 4 Drive"      "Hornet Sportabout"   "Valiant"
## [7] "Duster 360"          "Merc 240D"           "Merc 230"
## [10] "Merc 280"            "Merc 280C"           "Merc 450SE"
## [13] "Merc 450SL"          "Merc 450SLC"         "Cadillac Fleetwood"
## [16] "Lincoln Continental" "Chrysler Imperial"   "Fiat 128"
## [19] "Honda Civic"         "Toyota Corolla"      "Toyota Corona"
```

```
## [22] "Dodge Challenger"      "AMC Javelin"          "Camaro Z28"
## [25] "Pontiac Firebird"      "Fiat X1-9"            "Porsche 914-2"
## [28] "Lotus Europa"          "Ford Pantera L"       "Ferrari Dino"
## [31] "Maserati Bora"         "Volvo 142E"
```

```
mtcars$carb
```

```
## [1] 4 4 1 1 2 1 4 2 2 4 4 3 3 3 4 4 4 1 2 1 1 2 2 4 2 1 2 2 4 6 8 2
```

We can also view particular subsets of the data frame by using the `df[row, column]` notation.

```
mtcars[2, 4] # access the value from the second row, fourth column (this is the horsepower of a Mazda R
```

```
## [1] 110
```

```
mtcars[1, ] # access all values in the first row (all info on the Mazda RX4)
```

```
##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4  21   6  160 110  3.9 2.62 16.46  0  1    4    4
```

```
mtcars[, 1] # access all values in the first column (mpg info for all cars)
```

```
## [1] 21.0 21.0 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 17.8 16.4 17.3 15.2 10.4
## [16] 10.4 14.7 32.4 30.4 33.9 21.5 15.5 15.2 13.3 19.2 27.3 26.0 30.4 15.8 19.7
## [31] 15.0 21.4
```

```
mtcars[1:5, c(1, 4, 6)] # rows one through five and columns 1, 4, and 6
```

```
##           mpg  hp   wt
## Mazda RX4    21.0 110 2.620
## Mazda RX4 Wag 21.0 110 2.875
## Datsun 710    22.8  93 2.320
## Hornet 4 Drive 21.4 110 3.215
## Hornet Sportabout 18.7 175 3.440
```

```
mtcars[-(3:30), ] # access everything EXCEPT rows 3 through 30
```

```
##           mpg cyl disp  hp drat   wt  qsec vs am gear carb
## Mazda RX4    21.0   6  160 110  3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag 21.0   6  160 110  3.90 2.875 17.02  0  1    4    4
## Maserati Bora 15.0   8  301 335  3.54 3.570 14.60  0  1    5    8
## Volvo 142E    21.4   4  121 109  4.11 2.780 18.60  1  1    4    2
```

```
mtcars[, c(-(1:4), -6)] # access everything EXCEPT columns 1 through 4 and column 6
```

```
##           drat  qsec vs am gear carb
## Mazda RX4    3.90 16.46  0  1    4    4
## Mazda RX4 Wag 3.90 17.02  0  1    4    4
## Datsun 710    3.85 18.61  1  1    4    1
## Hornet 4 Drive 3.08 19.44  1  0    3    1
## Hornet Sportabout 3.15 17.02  0  0    3    2
## Valiant      2.76 20.22  1  0    3    1
## Duster 360   3.21 15.84  0  0    3    4
## Merc 240D    3.69 20.00  1  0    4    2
## Merc 230     3.92 22.90  1  0    4    2
## Merc 280     3.92 18.30  1  0    4    4
## Merc 280C    3.92 18.90  1  0    4    4
## Merc 450SE   3.07 17.40  0  0    3    3
## Merc 450SL   3.07 17.60  0  0    3    3
## Merc 450SLC  3.07 18.00  0  0    3    3
```

```
## Cadillac Fleetwood  2.93 17.98  0  0   3   4
## Lincoln Continental 3.00 17.82  0  0   3   4
## Chrysler Imperial   3.23 17.42  0  0   3   4
## Fiat 128             4.08 19.47  1  1   4   1
## Honda Civic          4.93 18.52  1  1   4   2
## Toyota Corolla       4.22 19.90  1  1   4   1
## Toyota Corona        3.70 20.01  1  0   3   1
## Dodge Challenger     2.76 16.87  0  0   3   2
## AMC Javelin          3.15 17.30  0  0   3   2
## Camaro Z28           3.73 15.41  0  0   3   4
## Pontiac Firebird     3.08 17.05  0  0   3   2
## Fiat X1-9            4.08 18.90  1  1   4   1
## Porsche 914-2        4.43 16.70  0  1   5   2
## Lotus Europa         3.77 16.90  1  1   5   2
## Ford Pantera L       4.22 14.50  0  1   5   4
## Ferrari Dino         3.62 15.50  0  1   5   6
## Maserati Bora        3.54 14.60  0  1   5   8
## Volvo 142E          4.11 18.60  1  1   4   2
```

```
mtcars["Duster 360" , "mpg"] # select the mpg (column) for only Duster 360 (row)
```

```
## [1] 14.3
```

Select the horse power data for the Duster 360, Lincoln Continental, and Maserati Bora. Instead of manually viewing the matrix to determine which rows and column numbers you want to specify, use `rownames()` and `colnames()`.

```
rows_of_interest <- rownames(mtcars) %in% c("Duster 360", "Lincoln Continental", "Maserati Bora")
hp_col_index <- which(colnames(mtcars) == "hp")
mtcars[rows_of_interest, hp_col_index]
```

```
## [1] 245 215 335
```

Oftentimes we want to edit data frames to add calculated columns, delete unnecessary rows, subset to a particular view, etc.

You can create a column using the “\$” notation to specify the new column name, and “<-” to assign some values to the new column. The syntax is as follows:

```
values <- c(5, 7, 12, 20)
mtcars$new_column <- values
```

Let’s try this out! In the `mtcars` data frame, there is a column called “am”. This column holds transmission information where 0 = automatic, and 1 = manual. But as of now, this isn’t very clear without knowing that information.

Create a new column called “transmission” that assigns the value “automatic” where `am = 0`, and “manual” where `am = 1`.

You can use an `ifelse` statement to assign your values! Remember, `ifelse()` takes in three arguments: `ifelse(test_condition, yes, no)`

```
mtcars$transmission <- ifelse(mtcars$am == 0, "automatic", "manual")
head(mtcars[, c("am", "transmission")])
```

```
##           am transmission
```

```
## Mazda RX4          1      manual
## Mazda RX4 Wag      1      manual
## Datsun 710          1      manual
## Hornet 4 Drive     0      automatic
## Hornet Sportabout  0      automatic
## Valiant            0      automatic
```

Now that we have the “transmission” column, the “am” column is no longer particularly useful. Let’s delete it.

You can easily delete a row/column by subsetting around it using `df[row, column]` notation. Simply choose all of the rows/columns you want to keep (ignoring the ones you want to delete), and assign the subset to either the original data frame variable (to permanently delete the rows/columns), or to a new variable storing an edited copy of the data frame. In general, it is best to create a copy that way you always have access to your original unedited data frame if needed.

The syntax is as follows:

```
# df_copy <- df[rows_to_keep, columns_to_keep]
```

Use the above syntax to save a copy of the `mtcars` data without the “am” column. Remember to use `colnames()` to find the column number of “am”. Also recall that to select row/columns to drop (hence selecting everything else), you can use “-index #”.

```
am_col_index <- which(colnames(mtcars) == "am")
mtcars_copy <- mtcars[, -am_col_index]
```

View your new `mtcars` copy data frame to check that the “am” column was dropped.

```
colnames(mtcars_copy)

## [1] "mpg"      "cyl"      "disp"      "hp"      "drat"
## [6] "wt"       "qsec"     "vs"       "gear"    "carb"
## [11] "new_column" "transmission"
```

(2) Query and retrieve the TCGA breast cancer clinical data

Now that you’ve worked with a simpler example dataset, it is time to work with TCGA breast cancer clinical dataset!

Hint: The code you use here is the same as the code when you initially queried, downloaded, and retrieved the TCGA clinical data.

query TCGA BRCA clinical data

```
library(TCGAbiolinks)

# query TCGA BRCA clinical data
clin_query <- "TCGA-BRCA"
clinic <- GDCQuery_clinic(project = clin_query, type = "clinical")

colnames(clinic)[colnames(clinic) == "bcr_patient_barcode"] <- "Tumor_Sample_Barcode"
```

Querying TCGA data can take a while, even without downloading. A faster way to save and access the clinical data is to save the dataframe as a csv file on your local computer. Use the following command to do so:

```
#write.csv(clinic, "brca_clinical_data.csv", row.names=FALSE)
```

To read the csv back into your environment as a dataframe:

```
#clinic <- read.csv("brca_clinical_data.csv")
```

(3) Exploring the Data

View the structure of the data as well as the first few lines. What two commands should you use? Answer the following questions based on the output of those commands: # 1. How many rows are there? How many columns? ## 1098 rows and 107 columns. # 2. What is the data type of \$gender? ## Data type is character # 3. What is the age in years of the 5th patient (barcode: TCGA-4H-AAAK)? ## The age in years of 5th patient is 50. # 4. How old was the 1st patient at their last follow up (barcode: TCGA-3C-AAAU)? HINT: you need to look at data from two columns to answer this ## The age of the 1st patient at their last follow up is 66.

```
str(clinic)
```

```
## 'data.frame': 1098 obs. of 103 variables:
## $ project : chr "TCGA-BRCA" "TCGA-BRCA" "TCGA-BRCA" "TCGA-BRCA" ...
## $ submitter_id : chr "TCGA-A7-A0DC" "TCGA-Z7-A8R6" "TCGA-C8-...
## $ synchronous_malignancy : chr "No" "No" "No" "No" ...
## $ ajcc_pathologic_stage : chr "Stage IA" "Stage I" "Stage IIA" "Stage IIB" ...
## $ days_to_diagnosis : int 0 0 0 0 0 0 0 0 0 0 ...
## $ laterality : chr NA "Left" "Right" "Left" ...
## $ created_datetime : chr NA NA NA NA ...
## $ last_known_disease_status : chr "not reported" NA NA NA ...
## $ tissue_or_organ_of_origin : chr "Breast, NOS" "Overlapping lesion of breast" "Breast, NOS" ...
## $ age_at_diagnosis : int 23294 16955 NA 25475 31805 24891 27163 ...
## $ primary_diagnosis : chr "Infiltrating duct carcinoma, NOS" "Plasma cell myeloma" "Infiltrating duct carcinoma, NOS" ...
## $ updated_datetime : chr "2025-01-08T12:58:05.350809-06:00" "2025-01-08T12:58:05.350809-06:00" ...
## $ prior_malignancy : chr "no" "no" "no" "no" ...
## $ year_of_diagnosis : int 2009 2005 2010 2007 2004 1998 2010 2009 ...
## $ state : chr "released" "released" "released" "released" ...
## $ prior_treatment : chr "No" "No" "No" "No" ...
## $ diagnosis_is_primary_disease : logi NA TRUE TRUE TRUE TRUE TRUE ...
## $ days_to_last_known_disease_status : logi NA NA NA NA NA NA ...
## $ method_of_diagnosis : chr NA "Surgical Resection" "Surgical Resection" "Surgical Resection" ...
## $ ajcc_staging_system_edition : chr NA "6th" "6th" "6th" ...
## $ ajcc_pathologic_t : chr "T1c" "T1c" "T2" "T2" ...
## $ days_to_recurrence : logi NA NA NA NA NA NA ...
## $ morphology : chr "8500/3" "8022/3" "8500/3" "8500/3" ...
## $ ajcc_pathologic_n : chr "N0 (i-)" "N0" "N0" "N0" ...
## $ ajcc_pathologic_m : chr "M0" "M0" "M0" "M0" ...
## $ classification_of_tumor : chr "not reported" "primary" "primary" "primary" ...
## $ diagnosis_id : chr "bfd786ee-b81c-5eb5-9da8-44e35c29baf9" "bfd786ee-b81c-5eb5-9da8-44e35c29baf9" ...
## $ icd_10_code : chr "C50.9" "C50.8" "C50.9" "C50.9" ...
## $ site_of_resection_or_biopsy : chr "Breast, NOS" "Breast, NOS" "Breast, NOS" "Breast, NOS" ...
## $ tumor_grade : chr "Not Reported" NA NA NA ...
## $ sites_of_involvement :List of 1098
## ..$ : NULL
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Left Lower Outer"
## ..$ : chr "Breast, Right Upper Inner"
## ..$ : chr "Breast, Right Lower Outer"
```

```

## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Right Upper Outer"
## ..$ : chr "Breast, Left Lower Inner"
## ..$ : chr "Breast, NOS" "Breast, Left Lower Inner"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Upper Outer" "Breast, NOS"
## ..$ : chr "Breast, Right Upper Inner"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Left Upper Inner" "Breast, Left Upper Outer"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Right Lower Outer"
## ..$ : chr "Breast, Left Lower Inner" "Breast, NOS"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Upper Inner"
## ..$ : chr "Breast, Right Upper Inner"
## ..$ : chr "Breast, NOS" "Breast, Right Upper Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Lower Inner"
## ..$ : chr "Breast, Right Upper Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Right Upper Inner"
## ..$ : chr "Breast, Left Lower Inner"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Right Upper Inner" "Breast, NOS"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Right Upper Outer"
## ..$ : chr "Breast, Right Lower Inner"
## ..$ : chr "Breast, Right Lower Outer"
## ..$ : chr "Breast, Right Upper Outer"
## ..$ : chr "Breast, Left Lower Inner"
## ..$ : chr "Breast, Right Lower Outer"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Right Upper Inner"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Right Lower Inner"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, NOS" "Breast, Left Upper Inner"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Left Lower Outer"
## ..$ : chr "Breast, Left Upper Inner"

```

```

## ..$ : chr "Breast, Right Upper Inner"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Right Upper Outer" "Breast, NOS"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Right Lower Outer"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Bone, NOS" "Breast, Left Lower Outer" "Breast, NOS"
## ..$ : chr "Bone, NOS" "Liver" "Breast, NOS" "Breast, Right Upper Inner" ...
## ..$ : chr "Breast, Left Upper Outer" "Breast, NOS"
## ..$ : chr "Breast, Left Lower Outer"
## ..$ : chr "Breast, Left Lower Inner"
## ..$ : chr "Breast, Left Upper Outer" "Breast, NOS"
## ..$ : chr "Breast, NOS" "Breast, Left Upper Outer" "Breast, Left Lower Outer"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, NOS" "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Lower Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Right Upper Inner"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Right Upper Outer" "Breast, NOS"
## ..$ : chr "Breast, Right Upper Outer" "Breast, NOS"
## ..$ : chr "Breast, Right Upper Inner"
## ..$ : chr "Breast, Left Lower Inner" "Breast, NOS"
## ..$ : chr "Breast, Right Upper Inner" "Breast, NOS"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Right Upper Outer"
## ..$ : chr "Breast, NOS"
## ..$ : chr "Breast, Right Upper Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, Left Upper Outer"
## ..$ : chr "Breast, NOS"
## .. [list output truncated]
## $ progression_or_recurrence : chr "not reported" NA NA NA ...
## $ metastasis_at_diagnosis : chr NA NA "No Metastasis" NA ...
## $ tumor_of_origin : chr NA NA NA NA ...
## $ NA. : logi NA NA NA NA NA NA ...
## $ figo_stage : chr NA NA NA NA ...
## $ figo_staging_edition_year : chr NA NA NA NA ...
## $ cigarettes_per_day : logi NA NA NA NA NA NA ...
## $ alcohol_history : chr "Not Reported" NA NA NA ...
## $ exposure_id : chr "cf0e2785-4dc2-5f31-9a68-d462e5295cdd"
## $ years_smoked : logi NA NA NA NA NA NA ...
## $ alcohol_intensity : logi NA NA NA NA NA NA ...
## $ race : chr "white" "white" "asian" "not reported"
## $ gender : chr "female" "female" "female" "female" ..
## $ ethnicity : chr "not hispanic or latino" "not hispanic"

```



```

## $ vital_status : chr "Alive" "Alive" "Alive" "Alive" ...
## $ age_at_index : int 63 46 59 69 87 68 74 70 72 59 ...
## $ days_to_birth : int -23294 -16955 NA -25475 -31805 -24891 ...
## $ year_of_birth : int 1946 NA NA NA NA NA NA NA NA ...
## $ demographic_id : chr "e4fddcc5-5609-591a-9ffb-835b28d3d7ac"
## $ age_is_obfuscated : logi NA FALSE FALSE FALSE FALSE FALSE ...
## $ year_of_death : logi NA NA NA NA NA NA ...
## $ country_of_residence_at_enrollment : chr NA "United States" "Vietnam" "Germany" ...
## $ days_to_death : int NA NA NA NA 792 NA NA NA NA NA ...
## $ days_to_last_follow_up : int NA 3256 375 1308 792 3088 10 608 611 5 ...
## $ follow_ups_disease_response : chr NA "TF-Tumor Free" "TF-Tumor Free" "TF-Tumor Free" ...
## $ treatments_pharmaceutical_days_to_treatment_end : chr "NA" "153" "NA" "NA" ...
## $ treatments_pharmaceutical_days_to_treatment_start : chr "NA" "31" "NA" "NA" ...
## $ treatments_pharmaceutical_treatment_id : chr "4c866bf4-0a1e-53d1-ae6-6716e4da409e"
## $ treatments_pharmaceutical_treatment_type : chr "Pharmaceutical Therapy, NOS" "Pharmaceutical Therapy, NOS" ...
## $ treatments_pharmaceutical_regimen_or_line_of_therapy : chr "NA" "NA" "NA" "NA" ...
## $ treatments_pharmaceutical_treatment_effect : chr "NA" "NA" "NA" "NA" ...
## $ treatments_pharmaceutical_therapeutic_agents : chr "NA" "Doxorubicin Hydrochloride;Cyclophosphamide" ...
## $ treatments_pharmaceutical_treatment_or_therapy : chr "yes" "yes" "yes" "no" ...
## $ treatments_pharmaceutical_initial_disease_status : chr "NA" "NA" "NA" "NA" ...
## $ treatments_pharmaceutical_treatment_intent_type : chr "NA" "NA" "Adjuvant" "Adjuvant" ...
## $ treatments_pharmaceutical_treatment_anatomic_site : chr "NA" "NA" "NA" "NA" ...
## $ treatments_pharmaceutical_treatment_outcome : chr "NA" "Complete Response" "Not Reported" ...
## $ treatments_pharmaceutical_treatment_anatomic_sites : chr "NULL" "NULL" "NULL" "NULL" ...
## $ treatments_pharmaceutical_margin_status : chr "NA" "NA" "NA" "NA" ...
## $ treatments_pharmaceutical_clinical_trial_indicator : chr "NA" "No" "NA" "NA" ...
## $ treatments_pharmaceutical_course_number : chr "NA" "NA" "NA" "NA" ...
## $ treatments_pharmaceutical_treatment_dose : chr "NA" "NA" "NA" "NA" ...
## $ treatments_pharmaceutical_number_of_fractions : chr "NA" "NA" "NA" "NA" ...
## $ treatments_pharmaceutical_treatment_dose_units : chr "NA" "NA" "NA" "NA" ...
## $ treatments_pharmaceutical_prescribed_dose_units : chr "NA" "NA" "NA" "NA" ...
## $ treatments_pharmaceutical_route_of_administration : chr "NULL" "NULL" "NULL" "NULL" ...
## $ treatments_pharmaceutical_prescribed_dose : chr "NA" "NA" "NA" "NA" ...
## $ treatments_pharmaceutical_number_of_cycles : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_days_to_treatment_end : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_days_to_treatment_start : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_treatment_id : chr "747d663f-737d-5abf-a40a-ed5ea2656f17"
## $ treatments_radiation_treatment_type : chr "Radiation Therapy, NOS" "Radiation Therapy, NOS" ...
## $ treatments_radiation_regimen_or_line_of_therapy : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_treatment_effect : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_therapeutic_agents : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_treatment_or_therapy : chr "yes" "no" "no" "no" ...
## $ treatments_radiation_initial_disease_status : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_treatment_intent_type : chr "NA" "Adjuvant" "Adjuvant" "Adjuvant" ...
## $ treatments_radiation_treatment_anatomic_site : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_treatment_outcome : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_treatment_anatomic_sites : chr "NULL" "NULL" "NULL" "NULL" ...
## $ treatments_radiation_margin_status : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_clinical_trial_indicator : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_course_number : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_treatment_dose : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_number_of_fractions : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_treatment_dose_units : chr "NA" "NA" "NA" "NA" ...
## $ treatments_radiation_prescribed_dose_units : chr "NA" "NA" "NA" "NA" ...

```

```
## [list output truncated]
```

```
head(clinic)
```

```
##      project submitter_id synchronous_malignancy ajcc_pathologic_stage
## 1 TCGA-BRCA TCGA-A7-A0DC                      No                Stage IA
## 2 TCGA-BRCA TCGA-Z7-A8R6                      No                Stage I
## 3 TCGA-BRCA TCGA-C8-A1HE                      No                Stage IIA
## 4 TCGA-BRCA TCGA-A8-A07B                      No                Stage IIA
## 5 TCGA-BRCA TCGA-AC-A2FM                      No                Stage IIB
## 6 TCGA-BRCA TCGA-B6-A1KF                      No                Stage IIB
##      days_to_diagnosis laterality created_datetime last_known_disease_status
## 1              0      <NA>          <NA>                not reported
## 2              0      Left          <NA>                <NA>
## 3              0     Right          <NA>                <NA>
## 4              0     Left          <NA>                <NA>
## 5              0     Right          <NA>                <NA>
## 6              0     Right          <NA>                <NA>
##      tissue_or_organ_of_origin age_at_diagnosis
## 1              Breast, NOS          23294
## 2 Overlapping lesion of breast      16955
## 3              Breast, NOS           NA
## 4              Breast, NOS          25475
## 5              Breast, NOS          31805
## 6              Breast, NOS          24891
##      primary_diagnosis          updated_datetime
## 1 Infiltrating duct carcinoma, NOS 2025-01-08T12:58:05.350809-06:00
## 2      Pleomorphic carcinoma 2025-03-26T09:15:23.441986-05:00
## 3 Infiltrating duct carcinoma, NOS 2025-03-26T09:15:23.441986-05:00
## 4 Infiltrating duct carcinoma, NOS 2025-03-26T09:15:23.441986-05:00
## 5      Lobular carcinoma, NOS 2025-03-26T09:15:23.441986-05:00
## 6 Infiltrating duct carcinoma, NOS 2025-03-26T09:15:23.441986-05:00
##      prior_malignancy year_of_diagnosis      state prior_treatment
## 1              no          2009 released          No
## 2              no          2005 released          No
## 3              no          2010 released          No
## 4              no          2007 released          No
## 5              no          2004 released          No
## 6              no          1998 released          No
##      diagnosis_is_primary_disease days_to_last_known_disease_status
## 1              NA                NA
## 2              TRUE                NA
## 3              TRUE                NA
## 4              TRUE                NA
## 5              TRUE                NA
## 6              TRUE                NA
##      method_of_diagnosis ajcc_staging_system_edition ajcc_pathologic_t
## 1              <NA>          <NA>                T1c
## 2      Surgical Resection          6th                T1c
## 3      Surgical Resection          6th                T2
## 4              <NA>          6th                T2
## 5      Core Biopsy          6th                T2
## 6 Fine Needle Aspiration          5th                T2
##      days_to_recurrence morphology ajcc_pathologic_n ajcc_pathologic_m
## 1              NA      8500/3          NO (i-)                M0
```

## 2	NA	8022/3	NO	MO	
## 3	NA	8500/3	NO	MO	
## 4	NA	8500/3	NO	MO	
## 5	NA	8520/3	N1mi	MO	
## 6	NA	8500/3	N1	MO	
##	classification_of_tumor		diagnosis_id	icd_10_code	
## 1	not reported	bfd786ee-b81c-5eb5-9da8-44e35c29baf9		C50.9	
## 2	primary	71037d87-bfdd-56d2-874f-fc4dd4fa9932		C50.8	
## 3	primary	65407607-8c12-53b8-81a2-6f84945a847a		C50.9	
## 4	primary	9f0c2231-31cc-56eb-9a7e-948350bd9d3c		C50.9	
## 5	primary	589b6257-1230-5d67-bf2a-7cfe3a00fd05		C50.9	
## 6	primary	886eae4a-ea85-579c-b1ac-52c55304f502		C50.9	
##	site_of_resection_or_biopsy	tumor_grade	sites_of_involvement		
## 1	Breast, NOS	Not Reported			
## 2	Breast, NOS	<NA>	Breast,		
## 3	Breast, NOS	<NA>	Breast, NOS		
## 4	Breast, NOS	<NA>	Breast,		
## 5	Breast, NOS	<NA>	Breast,		
## 6	Breast, NOS	<NA>	Breast,		
##	progression_or_recurrence	metastasis_at_diagnosis	tumor_of_origin	NA.	
## 1	not reported	<NA>	<NA>	NA	
## 2	<NA>	<NA>	<NA>	NA	
## 3	<NA>	No Metastasis	<NA>	NA	
## 4	<NA>	<NA>	<NA>	NA	
## 5	<NA>	Metastasis, NOS	<NA>	NA	
## 6	<NA>	No Metastasis	<NA>	NA	
##	figo_stage	figo_staging_edition_year	cigarettes_per_day	alcohol_history	
## 1	<NA>	<NA>	NA	Not Reported	
## 2	<NA>	<NA>	NA	<NA>	
## 3	<NA>	<NA>	NA	<NA>	
## 4	<NA>	<NA>	NA	<NA>	
## 5	<NA>	<NA>	NA	<NA>	
## 6	<NA>	<NA>	NA	<NA>	
##		exposure_id	years_smoked	alcohol_intensity	
## 1	cf0e2785-4dc2-5f31-9a68-d462e5295cdd	NA		NA	
## 2		<NA>	NA	NA	
## 3		<NA>	NA	NA	
## 4		<NA>	NA	NA	
## 5		<NA>	NA	NA	
## 6		<NA>	NA	NA	
##	race	gender	ethnicity	vital_status	age_at_index
## 1	white	female	not hispanic or latino	Alive	63
## 2	white	female	not hispanic or latino	Alive	46
## 3	asian	female	not hispanic or latino	Alive	59
## 4	not reported	female	not reported	Alive	69
## 5	white	female	not hispanic or latino	Dead	87
## 6	white	female	not hispanic or latino	Alive	68
##	days_to_birth	year_of_birth		demographic_id	
## 1	-23294	1946	e4fddcc5-5609-591a-9fffb-835b28d3d7ac		
## 2	-16955	NA	ed1a9ed4-1b8f-5394-ada9-84750a14d767		
## 3	NA	NA	58f39567-33fc-52c5-b323-a98a553425e7		
## 4	-25475	NA	193426d2-e48a-57b9-b3fa-be830685301b		
## 5	-31805	NA	78da5fb8-73d6-5966-be41-109fb625bf18		
## 6	-24891	NA	da42095a-0643-5e99-9e7e-2d5b6e8e1d8d		

	age_is_obfuscated	year_of_death	country_of_residence_at_enrollment
## 1	NA	NA	<NA>
## 2	FALSE	NA	United States
## 3	FALSE	NA	Vietnam
## 4	FALSE	NA	Germany
## 5	FALSE	NA	United States
## 6	FALSE	NA	United States

	days_to_death	days_to_last_follow_up	follow_ups_disease_response
## 1	NA	NA	<NA>
## 2	NA	3256	TF-Tumor Free
## 3	NA	375	TF-Tumor Free
## 4	NA	1308	TF-Tumor Free
## 5	792	792	WT-With Tumor
## 6	NA	3088	TF-Tumor Free

	treatments_pharmaceutical_days_to_treatment_end
## 1	NA
## 2	153
## 3	NA
## 4	NA
## 5	NA
## 6	NA

	treatments_pharmaceutical_days_to_treatment_start
## 1	NA
## 2	31
## 3	NA
## 4	NA
## 5	NA
## 6	NA

	treatments_pharmaceutical_id
## 1	4c866bf4-0a1e-53d1-ae66-000000000000
## 2	864e1425-595b-596e-92cb-f2e9a79e0aa0;8c066ce4-0f07-4636-9714-3293037e246c;fe427b01-d901-4c6c-893a-c00000000000
## 3	be81bca6-040a-5ab4-9837-c00000000000
## 4	f732ff5a-923d-5701-924c-c00000000000
## 5	135a30fc-2010-5b94-853f-500000000000
## 6	074ef8b8-2a6b-5cb3-9b2c-c00000000000

	treatments_pharmaceutical_treatment_type
## 1	Pharmaceutical Therapy, NOS
## 2	Pharmaceutical Therapy, NOS
## 3	Pharmaceutical Therapy, NOS
## 4	Pharmaceutical Therapy, NOS
## 5	Pharmaceutical Therapy, NOS
## 6	Pharmaceutical Therapy, NOS

	treatments_pharmaceutical_regimen_or_line_of_therapy
## 1	NA
## 2	NA
## 3	NA
## 4	NA
## 5	NA
## 6	NA

	treatments_pharmaceutical_treatment_effect
## 1	NA
## 2	NA
## 3	NA
## 4	NA

##	5	NA
##	6	NA
##		treatments_pharmaceutical_therapeutic_agents
##	1	NA
##	2	Doxorubicin Hydrochloride;Cyclophosphamide;Paclitaxel
##	3	NA
##	4	NA
##	5	NA
##	6	NA
##		treatments_pharmaceutical_treatment_or_therapy
##	1	yes
##	2	yes
##	3	yes
##	4	no
##	5	no
##	6	yes
##		treatments_pharmaceutical_initial_disease_status
##	1	NA
##	2	NA
##	3	NA
##	4	NA
##	5	NA
##	6	NA
##		treatments_pharmaceutical_treatment_intent_type
##	1	NA
##	2	NA
##	3	Adjuvant
##	4	Adjuvant
##	5	Adjuvant
##	6	Adjuvant
##		treatments_pharmaceutical_treatment_anatomic_site
##	1	NA
##	2	NA
##	3	NA
##	4	NA
##	5	NA
##	6	NA
##		treatments_pharmaceutical_treatment_outcome
##	1	NA
##	2	Complete Response
##	3	Not Reported
##	4	NA
##	5	NA
##	6	Not Reported
##		treatments_pharmaceutical_treatment_anatomic_sites
##	1	NULL
##	2	NULL
##	3	NULL
##	4	NULL
##	5	NULL
##	6	NULL
##		treatments_pharmaceutical_margin_status
##	1	NA
##	2	NA

## 3	NA
## 4	NA
## 5	NA
## 6	NA
## treatments_pharmaceutical_clinical_trial_indicator	
## 1	NA
## 2	No
## 3	NA
## 4	NA
## 5	NA
## 6	NA
## treatments_pharmaceutical_course_number	
## 1	NA
## 2	NA
## 3	NA
## 4	NA
## 5	NA
## 6	NA
## treatments_pharmaceutical_treatment_dose	
## 1	NA
## 2	NA
## 3	NA
## 4	NA
## 5	NA
## 6	NA
## treatments_pharmaceutical_number_of_fractions	
## 1	NA
## 2	NA
## 3	NA
## 4	NA
## 5	NA
## 6	NA
## treatments_pharmaceutical_treatment_dose_units	
## 1	NA
## 2	NA
## 3	NA
## 4	NA
## 5	NA
## 6	NA
## treatments_pharmaceutical_prescribed_dose_units	
## 1	NA
## 2	NA
## 3	NA
## 4	NA
## 5	NA
## 6	NA
## treatments_pharmaceutical_route_of_administration	
## 1	NULL
## 2	NULL
## 3	NULL
## 4	NULL
## 5	NULL
## 6	NULL
## treatments_pharmaceutical_prescribed_dose	

##	1	NA	
##	2	NA	
##	3	NA	
##	4	NA	
##	5	NA	
##	6	NA	
##	treatments_pharmaceutical_number_of_cycles		
##	1	NA	
##	2	NA	
##	3	NA	
##	4	NA	
##	5	NA	
##	6	NA	
##	treatments_radiation_days_to_treatment_end		
##	1	NA	
##	2	NA	
##	3	NA	
##	4	NA	
##	5	NA	
##	6	NA	
##	treatments_radiation_days_to_treatment_start		
##	1	NA	
##	2	NA	
##	3	NA	
##	4	NA	
##	5	NA	
##	6	NA	
##	treatments_radiation_treatment_id treatments_radiation_treatment_type		
##	1	747d663f-737d-5abf-a40a-ed5ea2656f17	Radiation Therapy, NOS
##	2	2a47c271-a472-4759-91c7-5366ab9035c6	Radiation Therapy, NOS
##	3	a569b05f-3a4a-4dd5-8171-f51dc2fa863b	Radiation Therapy, NOS
##	4	a83ff0dd-a494-4f3a-b8b1-fcfe7e05c485	Radiation Therapy, NOS
##	5	9e404a15-e776-4c2d-a685-77a049bb6f21	Radiation Therapy, NOS
##	6	71b77e0c-fded-4f41-8fb1-c5771c0c455d	Radiation Therapy, NOS
##	treatments_radiation_regimen_or_line_of_therapy		
##	1	NA	
##	2	NA	
##	3	NA	
##	4	NA	
##	5	NA	
##	6	NA	
##	treatments_radiation_treatment_effect treatments_radiation_therapeutic_agents		
##	1	NA	NA
##	2	NA	NA
##	3	NA	NA
##	4	NA	NA
##	5	NA	NA
##	6	NA	NA
##	treatments_radiation_treatment_or_therapy		
##	1	yes	
##	2	no	
##	3	no	
##	4	no	
##	5	no	

## 6	yes	
##	treatments_radiation_initial_disease_status	
## 1	NA	
## 2	NA	
## 3	NA	
## 4	NA	
## 5	NA	
## 6	NA	
##	treatments_radiation_treatment_intent_type	
## 1	NA	
## 2	Adjuvant	
## 3	Adjuvant	
## 4	Adjuvant	
## 5	Adjuvant	
## 6	Adjuvant	
##	treatments_radiation_treatment_anatomic_site	
## 1	NA	
## 2	NA	
## 3	NA	
## 4	NA	
## 5	NA	
## 6	NA	
##	treatments_radiation_treatment_outcome	
## 1	NA	
## 2	NA	
## 3	NA	
## 4	NA	
## 5	NA	
## 6	Not Reported	
##	treatments_radiation_treatment_anatomic_sites	
## 1	NULL	
## 2	NULL	
## 3	NULL	
## 4	NULL	
## 5	NULL	
## 6	NULL	
##	treatments_radiation_margin_status	
## 1	NA	
## 2	NA	
## 3	NA	
## 4	NA	
## 5	NA	
## 6	NA	
##	treatments_radiation_clinical_trial_indicator	
## 1	NA	
## 2	NA	
## 3	NA	
## 4	NA	
## 5	NA	
## 6	NA	
##	treatments_radiation_course_number	treatments_radiation_treatment_dose
## 1	NA	NA
## 2	NA	NA
## 3	NA	NA


```

## 4 NA NA
## 5 NA NA
## 6 NA NA
## treatments_radiation_number_of_fractions
## 1 NA
## 2 NA
## 3 NA
## 4 NA
## 5 NA
## 6 NA
## treatments_radiation_treatment_dose_units
## 1 NA
## 2 NA
## 3 NA
## 4 NA
## 5 NA
## 6 NA
## treatments_radiation_prescribed_dose_units
## 1 NA
## 2 NA
## 3 NA
## 4 NA
## 5 NA
## 6 NA
## treatments_radiation_route_of_administration
## 1 NULL
## 2 NULL
## 3 NULL
## 4 NULL
## 5 NULL
## 6 NULL
## treatments_radiation_prescribed_dose treatments_radiation_number_of_cycles
## 1 NA NA
## 2 NA NA
## 3 NA NA
## 4 NA NA
## 5 NA NA
## 6 NA NA
## Tumor_Sample_Barcode
## 1 TCGA-A7-A0DC
## 2 TCGA-Z7-A8R6
## 3 TCGA-C8-A1HE
## 4 TCGA-A8-A07B
## 5 TCGA-AC-A2FM
## 6 TCGA-B6-A1KF

```

```
# 1. How many rows and columns are there?
```

```
dim(clinic) # returns (rows, columns)
```

```
## [1] 1098 103
```

```
# 2. What is the data type of $gender?
```

```
class(clinic$gender)
```

```
## [1] "character"
```

```
# 3. What is the age in years of the 5th patient (barcode: TCGA-4H-AAAK)?
clinic$age_years <- as.numeric(clinic$age_at_diagnosis) / 365.25
clinic[clinic$Tumor_Sample_Barcode == "TCGA-4H-AAAK", "age_years"]
```

```
## [1] 50.29706
```

```
# 4. How old was the 1st patient at their last follow up (barcode: TCGA-3C-AAAU)?
# need age_at_diagnosis + days_to_last_follow_up
patient1 <- clinic[clinic$Tumor_Sample_Barcode == "TCGA-3C-AAAU", ]
(patient1$age_at_diagnosis + patient1$days_to_last_follow_up) / 365.25
```

```
## [1] 66.41478
```

View the column and row names. What two commands should you use?

```
colnames(clinic)
```

```
## [1] "project"
## [2] "submitter_id"
## [3] "synchronous_malignancy"
## [4] "ajcc_pathologic_stage"
## [5] "days_to_diagnosis"
## [6] "laterality"
## [7] "created_datetime"
## [8] "last_known_disease_status"
## [9] "tissue_or_organ_of_origin"
## [10] "age_at_diagnosis"
## [11] "primary_diagnosis"
## [12] "updated_datetime"
## [13] "prior_malignancy"
## [14] "year_of_diagnosis"
## [15] "state"
## [16] "prior_treatment"
## [17] "diagnosis_is_primary_disease"
## [18] "days_to_last_known_disease_status"
## [19] "method_of_diagnosis"
## [20] "ajcc_staging_system_edition"
## [21] "ajcc_pathologic_t"
## [22] "days_to_recurrence"
## [23] "morphology"
## [24] "ajcc_pathologic_n"
## [25] "ajcc_pathologic_m"
## [26] "classification_of_tumor"
## [27] "diagnosis_id"
## [28] "icd_10_code"
## [29] "site_of_resection_or_biopsy"
## [30] "tumor_grade"
## [31] "sites_of_involvement"
## [32] "progression_or_recurrence"
## [33] "metastasis_at_diagnosis"
## [34] "tumor_of_origin"
## [35] "NA."
## [36] "figo_stage"
## [37] "figo_staging_edition_year"
## [38] "cigarettes_per_day"
## [39] "alcohol_history"
```

```

## [40] "exposure_id"
## [41] "years_smoked"
## [42] "alcohol_intensity"
## [43] "race"
## [44] "gender"
## [45] "ethnicity"
## [46] "vital_status"
## [47] "age_at_index"
## [48] "days_to_birth"
## [49] "year_of_birth"
## [50] "demographic_id"
## [51] "age_is_obfuscated"
## [52] "year_of_death"
## [53] "country_of_residence_at_enrollment"
## [54] "days_to_death"
## [55] "days_to_last_follow_up"
## [56] "follow_ups_disease_response"
## [57] "treatments_pharmaceutical_days_to_treatment_end"
## [58] "treatments_pharmaceutical_days_to_treatment_start"
## [59] "treatments_pharmaceutical_treatment_id"
## [60] "treatments_pharmaceutical_treatment_type"
## [61] "treatments_pharmaceutical_regimen_or_line_of_therapy"
## [62] "treatments_pharmaceutical_treatment_effect"
## [63] "treatments_pharmaceutical_therapeutic_agents"
## [64] "treatments_pharmaceutical_treatment_or_therapy"
## [65] "treatments_pharmaceutical_initial_disease_status"
## [66] "treatments_pharmaceutical_treatment_intent_type"
## [67] "treatments_pharmaceutical_treatment_anatomic_site"
## [68] "treatments_pharmaceutical_treatment_outcome"
## [69] "treatments_pharmaceutical_treatment_anatomic_sites"
## [70] "treatments_pharmaceutical_margin_status"
## [71] "treatments_pharmaceutical_clinical_trial_indicator"
## [72] "treatments_pharmaceutical_course_number"
## [73] "treatments_pharmaceutical_treatment_dose"
## [74] "treatments_pharmaceutical_number_of_fractions"
## [75] "treatments_pharmaceutical_treatment_dose_units"
## [76] "treatments_pharmaceutical_prescribed_dose_units"
## [77] "treatments_pharmaceutical_route_of_administration"
## [78] "treatments_pharmaceutical_prescribed_dose"
## [79] "treatments_pharmaceutical_number_of_cycles"
## [80] "treatments_radiation_days_to_treatment_end"
## [81] "treatments_radiation_days_to_treatment_start"
## [82] "treatments_radiation_treatment_id"
## [83] "treatments_radiation_treatment_type"
## [84] "treatments_radiation_regimen_or_line_of_therapy"
## [85] "treatments_radiation_treatment_effect"
## [86] "treatments_radiation_therapeutic_agents"
## [87] "treatments_radiation_treatment_or_therapy"
## [88] "treatments_radiation_initial_disease_status"
## [89] "treatments_radiation_treatment_intent_type"
## [90] "treatments_radiation_treatment_anatomic_site"
## [91] "treatments_radiation_treatment_outcome"
## [92] "treatments_radiation_treatment_anatomic_sites"
## [93] "treatments_radiation_margin_status"

```

```
## [94] "treatments_radiation_clinical_trial_indicator"
## [95] "treatments_radiation_course_number"
## [96] "treatments_radiation_treatment_dose"
## [97] "treatments_radiation_number_of_fractions"
## [98] "treatments_radiation_treatment_dose_units"
## [99] "treatments_radiation_prescribed_dose_units"
## [100] "treatments_radiation_route_of_administration"
## [101] "treatments_radiation_prescribed_dose"
## [102] "treatments_radiation_number_of_cycles"
## [103] "Tumor_Sample_Barcode"
## [104] "age_years"
```

```
rownames(clinic)
```

```
##      [1] "1"      "2"      "3"      "4"      "5"      "6"      "7"      "8"      "9"      "10"
##     [11] "11"     "12"     "13"     "14"     "15"     "16"     "17"     "18"     "19"     "20"
##     [21] "21"     "22"     "23"     "24"     "25"     "26"     "27"     "28"     "29"     "30"
##     [31] "31"     "32"     "33"     "34"     "35"     "36"     "37"     "38"     "39"     "40"
##     [41] "41"     "42"     "43"     "44"     "45"     "46"     "47"     "48"     "49"     "50"
##     [51] "51"     "52"     "53"     "54"     "55"     "56"     "57"     "58"     "59"     "60"
##     [61] "61"     "62"     "63"     "64"     "65"     "66"     "67"     "68"     "69"     "70"
##     [71] "71"     "72"     "73"     "74"     "75"     "76"     "77"     "78"     "79"     "80"
##     [81] "81"     "82"     "83"     "84"     "85"     "86"     "87"     "88"     "89"     "90"
##     [91] "91"     "92"     "93"     "94"     "95"     "96"     "97"     "98"     "99"     "100"
##    [101] "101"    "102"    "103"    "104"    "105"    "106"    "107"    "108"    "109"    "110"
##    [111] "111"    "112"    "113"    "114"    "115"    "116"    "117"    "118"    "119"    "120"
##    [121] "121"    "122"    "123"    "124"    "125"    "126"    "127"    "128"    "129"    "130"
##    [131] "131"    "132"    "133"    "134"    "135"    "136"    "137"    "138"    "139"    "140"
##    [141] "141"    "142"    "143"    "144"    "145"    "146"    "147"    "148"    "149"    "150"
##    [151] "151"    "152"    "153"    "154"    "155"    "156"    "157"    "158"    "159"    "160"
##    [161] "161"    "162"    "163"    "164"    "165"    "166"    "167"    "168"    "169"    "170"
##    [171] "171"    "172"    "173"    "174"    "175"    "176"    "177"    "178"    "179"    "180"
##    [181] "181"    "182"    "183"    "184"    "185"    "186"    "187"    "188"    "189"    "190"
##    [191] "191"    "192"    "193"    "194"    "195"    "196"    "197"    "198"    "199"    "200"
##    [201] "201"    "202"    "203"    "204"    "205"    "206"    "207"    "208"    "209"    "210"
##    [211] "211"    "212"    "213"    "214"    "215"    "216"    "217"    "218"    "219"    "220"
##    [221] "221"    "222"    "223"    "224"    "225"    "226"    "227"    "228"    "229"    "230"
##    [231] "231"    "232"    "233"    "234"    "235"    "236"    "237"    "238"    "239"    "240"
##    [241] "241"    "242"    "243"    "244"    "245"    "246"    "247"    "248"    "249"    "250"
##    [251] "251"    "252"    "253"    "254"    "255"    "256"    "257"    "258"    "259"    "260"
##    [261] "261"    "262"    "263"    "264"    "265"    "266"    "267"    "268"    "269"    "270"
##    [271] "271"    "272"    "273"    "274"    "275"    "276"    "277"    "278"    "279"    "280"
##    [281] "281"    "282"    "283"    "284"    "285"    "286"    "287"    "288"    "289"    "290"
##    [291] "291"    "292"    "293"    "294"    "295"    "296"    "297"    "298"    "299"    "300"
##    [301] "301"    "302"    "303"    "304"    "305"    "306"    "307"    "308"    "309"    "310"
##    [311] "311"    "312"    "313"    "314"    "315"    "316"    "317"    "318"    "319"    "320"
##    [321] "321"    "322"    "323"    "324"    "325"    "326"    "327"    "328"    "329"    "330"
##    [331] "331"    "332"    "333"    "334"    "335"    "336"    "337"    "338"    "339"    "340"
##    [341] "341"    "342"    "343"    "344"    "345"    "346"    "347"    "348"    "349"    "350"
##    [351] "351"    "352"    "353"    "354"    "355"    "356"    "357"    "358"    "359"    "360"
##    [361] "361"    "362"    "363"    "364"    "365"    "366"    "367"    "368"    "369"    "370"
##    [371] "371"    "372"    "373"    "374"    "375"    "376"    "377"    "378"    "379"    "380"
##    [381] "381"    "382"    "383"    "384"    "385"    "386"    "387"    "388"    "389"    "390"
##    [391] "391"    "392"    "393"    "394"    "395"    "396"    "397"    "398"    "399"    "400"
##    [401] "401"    "402"    "403"    "404"    "405"    "406"    "407"    "408"    "409"    "410"
```

##	[411]	"411"	"412"	"413"	"414"	"415"	"416"	"417"	"418"	"419"	"420"
##	[421]	"421"	"422"	"423"	"424"	"425"	"426"	"427"	"428"	"429"	"430"
##	[431]	"431"	"432"	"433"	"434"	"435"	"436"	"437"	"438"	"439"	"440"
##	[441]	"441"	"442"	"443"	"444"	"445"	"446"	"447"	"448"	"449"	"450"
##	[451]	"451"	"452"	"453"	"454"	"455"	"456"	"457"	"458"	"459"	"460"
##	[461]	"461"	"462"	"463"	"464"	"465"	"466"	"467"	"468"	"469"	"470"
##	[471]	"471"	"472"	"473"	"474"	"475"	"476"	"477"	"478"	"479"	"480"
##	[481]	"481"	"482"	"483"	"484"	"485"	"486"	"487"	"488"	"489"	"490"
##	[491]	"491"	"492"	"493"	"494"	"495"	"496"	"497"	"498"	"499"	"500"
##	[501]	"501"	"502"	"503"	"504"	"505"	"506"	"507"	"508"	"509"	"510"
##	[511]	"511"	"512"	"513"	"514"	"515"	"516"	"517"	"518"	"519"	"520"
##	[521]	"521"	"522"	"523"	"524"	"525"	"526"	"527"	"528"	"529"	"530"
##	[531]	"531"	"532"	"533"	"534"	"535"	"536"	"537"	"538"	"539"	"540"
##	[541]	"541"	"542"	"543"	"544"	"545"	"546"	"547"	"548"	"549"	"550"
##	[551]	"551"	"552"	"553"	"554"	"555"	"556"	"557"	"558"	"559"	"560"
##	[561]	"561"	"562"	"563"	"564"	"565"	"566"	"567"	"568"	"569"	"570"
##	[571]	"571"	"572"	"573"	"574"	"575"	"576"	"577"	"578"	"579"	"580"
##	[581]	"581"	"582"	"583"	"584"	"585"	"586"	"587"	"588"	"589"	"590"
##	[591]	"591"	"592"	"593"	"594"	"595"	"596"	"597"	"598"	"599"	"600"
##	[601]	"601"	"602"	"603"	"604"	"605"	"606"	"607"	"608"	"609"	"610"
##	[611]	"611"	"612"	"613"	"614"	"615"	"616"	"617"	"618"	"619"	"620"
##	[621]	"621"	"622"	"623"	"624"	"625"	"626"	"627"	"628"	"629"	"630"
##	[631]	"631"	"632"	"633"	"634"	"635"	"636"	"637"	"638"	"639"	"640"
##	[641]	"641"	"642"	"643"	"644"	"645"	"646"	"647"	"648"	"649"	"650"
##	[651]	"651"	"652"	"653"	"654"	"655"	"656"	"657"	"658"	"659"	"660"
##	[661]	"661"	"662"	"663"	"664"	"665"	"666"	"667"	"668"	"669"	"670"
##	[671]	"671"	"672"	"673"	"674"	"675"	"676"	"677"	"678"	"679"	"680"
##	[681]	"681"	"682"	"683"	"684"	"685"	"686"	"687"	"688"	"689"	"690"
##	[691]	"691"	"692"	"693"	"694"	"695"	"696"	"697"	"698"	"699"	"700"
##	[701]	"701"	"702"	"703"	"704"	"705"	"706"	"707"	"708"	"709"	"710"
##	[711]	"711"	"712"	"713"	"714"	"715"	"716"	"717"	"718"	"719"	"720"
##	[721]	"721"	"722"	"723"	"724"	"725"	"726"	"727"	"728"	"729"	"730"
##	[731]	"731"	"732"	"733"	"734"	"735"	"736"	"737"	"738"	"739"	"740"
##	[741]	"741"	"742"	"743"	"744"	"745"	"746"	"747"	"748"	"749"	"750"
##	[751]	"751"	"752"	"753"	"754"	"755"	"756"	"757"	"758"	"759"	"760"
##	[761]	"761"	"762"	"763"	"764"	"765"	"766"	"767"	"768"	"769"	"770"
##	[771]	"771"	"772"	"773"	"774"	"775"	"776"	"777"	"778"	"779"	"780"
##	[781]	"781"	"782"	"783"	"784"	"785"	"786"	"787"	"788"	"789"	"790"
##	[791]	"791"	"792"	"793"	"794"	"795"	"796"	"797"	"798"	"799"	"800"
##	[801]	"801"	"802"	"803"	"804"	"805"	"806"	"807"	"808"	"809"	"810"
##	[811]	"811"	"812"	"813"	"814"	"815"	"816"	"817"	"818"	"819"	"820"
##	[821]	"821"	"822"	"823"	"824"	"825"	"826"	"827"	"828"	"829"	"830"
##	[831]	"831"	"832"	"833"	"834"	"835"	"836"	"837"	"838"	"839"	"840"
##	[841]	"841"	"842"	"843"	"844"	"845"	"846"	"847"	"848"	"849"	"850"
##	[851]	"851"	"852"	"853"	"854"	"855"	"856"	"857"	"858"	"859"	"860"
##	[861]	"861"	"862"	"863"	"864"	"865"	"866"	"867"	"868"	"869"	"870"
##	[871]	"871"	"872"	"873"	"874"	"875"	"876"	"877"	"878"	"879"	"880"
##	[881]	"881"	"882"	"883"	"884"	"885"	"886"	"887"	"888"	"889"	"890"
##	[891]	"891"	"892"	"893"	"894"	"895"	"896"	"897"	"898"	"899"	"900"
##	[901]	"901"	"902"	"903"	"904"	"905"	"906"	"907"	"908"	"909"	"910"
##	[911]	"911"	"912"	"913"	"914"	"915"	"916"	"917"	"918"	"919"	"920"
##	[921]	"921"	"922"	"923"	"924"	"925"	"926"	"927"	"928"	"929"	"930"
##	[931]	"931"	"932"	"933"	"934"	"935"	"936"	"937"	"938"	"939"	"940"
##	[941]	"941"	"942"	"943"	"944"	"945"	"946"	"947"	"948"	"949"	"950"

```
## [951] "951" "952" "953" "954" "955" "956" "957" "958" "959" "960"
## [961] "961" "962" "963" "964" "965" "966" "967" "968" "969" "970"
## [971] "971" "972" "973" "974" "975" "976" "977" "978" "979" "980"
## [981] "981" "982" "983" "984" "985" "986" "987" "988" "989" "990"
## [991] "991" "992" "993" "994" "995" "996" "997" "998" "999" "1000"
## [1001] "1001" "1002" "1003" "1004" "1005" "1006" "1007" "1008" "1009" "1010"
## [1011] "1011" "1012" "1013" "1014" "1015" "1016" "1017" "1018" "1019" "1020"
## [1021] "1021" "1022" "1023" "1024" "1025" "1026" "1027" "1028" "1029" "1030"
## [1031] "1031" "1032" "1033" "1034" "1035" "1036" "1037" "1038" "1039" "1040"
## [1041] "1041" "1042" "1043" "1044" "1045" "1046" "1047" "1048" "1049" "1050"
## [1051] "1051" "1052" "1053" "1054" "1055" "1056" "1057" "1058" "1059" "1060"
## [1061] "1061" "1062" "1063" "1064" "1065" "1066" "1067" "1068" "1069" "1070"
## [1071] "1071" "1072" "1073" "1074" "1075" "1076" "1077" "1078" "1079" "1080"
## [1081] "1081" "1082" "1083" "1084" "1085" "1086" "1087" "1088" "1089" "1090"
## [1091] "1091" "1092" "1093" "1094" "1095" "1096" "1097" "1098"
```

View the vital status column. What two notations can be used to view columns?

```
clinic$vital_status
```

```
## [1] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive"
## [10] "Alive" "Dead" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive"
## [19] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [28] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [37] "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Dead" "Alive" "Alive"
## [46] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [55] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [64] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Dead"
## [73] "Dead" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive"
## [82] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead"
## [91] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive"
## [100] "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [109] "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead"
## [118] "Alive" "Alive" "Alive" "Dead" "Dead" "Alive" "Alive" "Alive" "Alive" "Dead"
## [127] "Alive" "Dead" "Alive" "Alive" "Dead" "Dead" "Alive" "Alive" "Alive" "Alive"
## [136] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead"
## [145] "Alive" "Alive" "Dead" "Alive" "Dead" "Alive" "Dead" "Alive" "Alive" "Dead"
## [154] "Alive" "Alive" "Dead" "Dead" "Alive" "Dead" "Alive" "Dead" "Dead" "Alive"
## [163] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive"
## [172] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive"
## [181] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive"
## [190] "Dead" "Dead" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive"
## [199] "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [208] "Dead" "Alive" "Alive" "Alive" "Alive" "Dead" "Dead" "Alive" "Alive"
## [217] "Dead" "Dead" "Dead" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive"
## [226] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [235] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Dead" "Alive"
## [244] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive"
## [253] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [262] "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [271] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [280] "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead"
## [289] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [298] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [307] "Dead" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Dead" "Alive"
```

[illegible]

```

## [802] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive"
## [811] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Dead" "Alive"
## [820] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [829] "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive"
## [838] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Dead" "Alive" "Dead"
## [847] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [856] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [865] "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive"
## [874] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [883] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [892] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [901] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [910] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [919] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [928] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive"
## [937] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [946] "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [955] "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive"
## [964] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive"
## [973] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [982] "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [991] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [1000] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [1009] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [1018] "Alive" "Dead" "Alive" "Alive" "Dead" "Alive" "Dead" "Dead" "Dead"
## [1027] "Dead" "Dead" "Alive" "Dead" "Alive" "Alive" "Dead" "Dead" "Alive"
## [1036] "Alive" "Dead" "Dead" "Dead" NA "Dead" "Dead" "Dead" "Dead"
## [1045] "Dead" "Dead" "Alive" "Alive" "Dead" "Alive" "Dead" "Dead" "Alive"
## [1054] "Alive" "Dead" "Alive" "Dead" "Alive" "Alive" "Alive" "Dead" "Dead"
## [1063] "Dead" "Dead" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Dead"
## [1072] "Dead" "Dead" "Alive" "Alive" "Alive" "Dead" "Alive" "Dead" "Alive"
## [1081] "Alive" "Dead" "Dead" "Dead" "Dead" "Alive" "Dead" "Dead" "Alive"
## [1090] "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive"

```

```
clinic[, "vital_status"]
```

```

## [1] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive"
## [10] "Alive" "Dead" "Dead" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive"
## [19] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [28] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [37] "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Dead" "Alive" "Alive"
## [46] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [55] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [64] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Dead"
## [73] "Dead" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive"
## [82] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead"
## [91] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive"
## [100] "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [109] "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead"
## [118] "Alive" "Alive" "Alive" "Dead" "Dead" "Alive" "Alive" "Alive" "Dead"
## [127] "Alive" "Dead" "Alive" "Alive" "Dead" "Dead" "Alive" "Alive" "Alive"
## [136] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead"
## [145] "Alive" "Alive" "Dead" "Alive" "Dead" "Alive" "Dead" "Alive" "Dead"
## [154] "Alive" "Alive" "Dead" "Dead" "Alive" "Dead" "Alive" "Dead" "Alive"
## [163] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive"

```


[illegible]

```

## [658] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [667] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [676] "Dead" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [685] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive"
## [694] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [703] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [712] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive"
## [721] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [730] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [739] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Dead" "Alive" "Alive" "Alive"
## [748] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [757] "Dead" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive"
## [766] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive"
## [775] "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive"
## [784] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [793] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [802] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive"
## [811] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Dead" "Alive" "Alive"
## [820] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [829] "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive"
## [838] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Dead" "Alive" "Dead" "Dead"
## [847] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [856] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [865] "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [874] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [883] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [892] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [901] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [910] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [919] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [928] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive"
## [937] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [946] "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [955] "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive"
## [964] "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive"
## [973] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [982] "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [991] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [1000] "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [1009] "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive" "Alive"
## [1018] "Alive" "Dead" "Alive" "Alive" "Dead" "Alive" "Dead" "Dead" "Dead" "Dead"
## [1027] "Dead" "Dead" "Alive" "Dead" "Alive" "Alive" "Dead" "Dead" "Dead" "Alive"
## [1036] "Alive" "Dead" "Dead" "Dead" NA "Dead" "Dead" "Dead" "Dead"
## [1045] "Dead" "Dead" "Alive" "Alive" "Dead" "Alive" "Dead" "Dead" "Dead" "Alive"
## [1054] "Alive" "Dead" "Alive" "Dead" "Alive" "Alive" "Alive" "Dead" "Dead" "Dead"
## [1063] "Dead" "Dead" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive" "Alive" "Dead"
## [1072] "Dead" "Dead" "Alive" "Alive" "Alive" "Dead" "Alive" "Dead" "Dead" "Alive"
## [1081] "Alive" "Dead" "Dead" "Dead" "Dead" "Alive" "Dead" "Dead" "Dead" "Alive"
## [1090] "Alive" "Alive" "Alive" "Alive" "Alive" "Dead" "Alive" "Alive" "Alive" "Alive"

```

(4) Plotting

Choose two variables and make a scatter plot (use the `plot()` function).

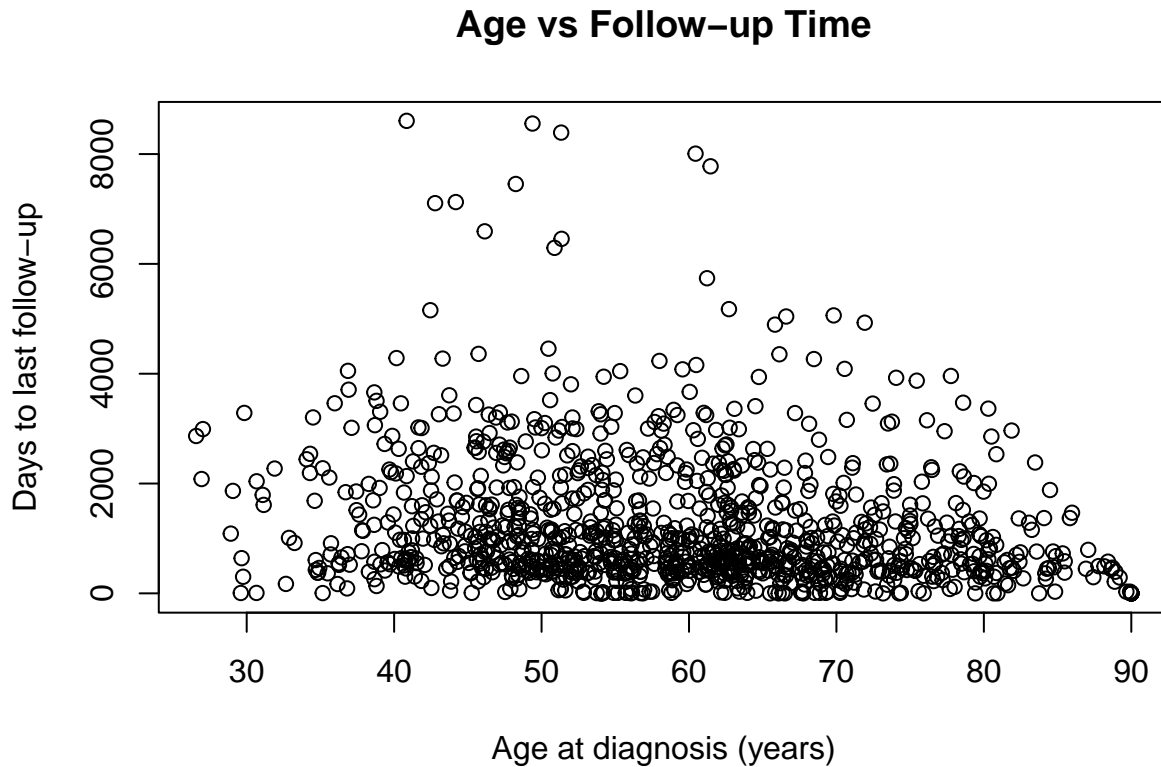
THINK: what types of variables can you use in a scatter plot (continuous, discrete, categorical). Look these

up if you need help!

For an extra challenge, add a plot title, axes titles, and colors to your plot.

```
clinic$age_years <- as.numeric(clinic$age_at_diagnosis) / 365.25
clinic$days_last_follow_up <- as.numeric(clinic$days_to_last_follow_up)

plot(clinic$age_years, clinic$days_last_follow_up,
     main = "Age vs Follow-up Time",
     xlab = "Age at diagnosis (years)",
     ylab = "Days to last follow-up")
```



Choose two new variables and make a box plot (use the `boxplot()` function).

THINK: what types of variables can you use in a box plot?

For an extra challenge, add a plot title, axes titles, and confidence intervals to your plot.

```
clinic$days_to_death_num <- as.numeric(clinic$days_to_death)
clinic$survival_time <- ifelse(!is.na(clinic$days_to_death_num),
                              clinic$days_to_death_num,
                              clinic$days_last_follow_up)

boxplot(survival_time ~ gender, data = clinic,
       main = "Survival Time by Gender",
       xlab = "Gender",
       ylab = "Days")
```

Survival Time by Gender



(5) Boolean Masking

As discussed in lecture, a boolean mask is a vector of True and False values that allow for fast and efficient indexing of large dataframes in R. Boolean masks work as an extremely helpful implementation of R's vectorization abilities.

A significant portion of your coding work will be data cleaning and preparation using boolean masks. As such, it's important you are comfortable with making them.

- 1) Create a mask to remove all NA values from vital_status

```
# Filled in example for you!
vital_status_na_mask <- !is.na(clinic$vital_status)
```

- 2) Create a mask to select for male patients

```
male_mask <- clinic$gender == "male"
```

- 3) Create a mask to select for patients under the age of 35 OR over 60 at diagnosis.

```
clinic$age_years <- as.numeric(clinic$age_at_diagnosis) / 365.25
under30_over60_mask <- clinic$age_years < 35 | clinic$age_years > 60
```

- 4) Using your under30_over60_mask, create a new dataframe called no_middle_clinic which contains all the clinical data for patients under the age of 30 or over 60 at diagnosis

```
no_middle_clinic <- clinic[under30_over60_mask, ]
```

(6) Kaplan-Meier Plotting

To create a KM plot in R, we first need to load some required packages.

```
if (!require(survival)) {
  install.packages("survival")
}
```

```
## Loading required package: survival
if (!require(survminer)) {
  install.packages("survminer")
}

## Loading required package: survminer
## Loading required package: ggplot2
## Loading required package: ggpubr

##
## Attaching package: 'survminer'

## The following object is masked from 'package:survival':
##
##     myeloma

if (!require(ggplot2)) {
  install.packages("ggplot2")
}
```

After installing packages as needed, we still need to load them into R to use them in our code.

```
library(survival)
library(survminer)
library(ggplot2)
```

A KM plot allows us to stratify survival by some clinical feature (or -omic feature!) of interest, therefore providing a simple and understandable plot for seeing how different aspects of a patient's health influence the severity of a disease. Because this requires grouping by a feature, KM plots always involve some categorical value plotted against survival percentage over time. As such, we need to first choose or otherwise create a categorical variable of interest.

For this example, let's create a KM plot to look at the relationship between breast cancer patient age and their rate of survival.

1. Data Preprocessing

First, check if there are any NA values in `age_at_diagnosis` (Hint: use the `is.na()` function. If you don't remember how the function works, run `?is.na()` in the console). Use Boolean indexing to remove the NA values.

```
cleaned_clinic <- clinic

age_NA_mask <- is.na(cleaned_clinic$age_at_diagnosis)
cleaned_clinic <- cleaned_clinic[!age_NA_mask, ]
```

Create a new column called "age_category" in `cleaned_clinic` that contains the value "Young" if `age_at_diagnosis` is ≤ 35 , "Middle" if > 35 and < 50 , and "Old" if ≥ 50 . (Hint: use `ifelse()`)

Note: These age threshold are relatively arbitrary, and just the ages that we are using in this example.

```
cleaned_clinic$age_in_years <- cleaned_clinic$age_at_diagnosis / 365

cleaned_clinic$age_category <- ifelse(cleaned_clinic$age_in_years <= 35, "Young",
  ifelse(cleaned_clinic$age_in_years < 50, "Middle", "Old"))
```

KM Plots also requires a column for total number of days survived post diagnosis. However, TCGA records survival over time two ways: `last_contact_days_to` for patients who are still alive, and `death_days_to` for patients who passed away. If a patient is still alive, their value in the `death_days_to` column is the string “[Not Applicable]”. Therefore, we need to create a new column called `survival_time` that combines the information in these two columns so there are no NA values.

```
cleaned_clinic$survival_time <- ifelse(cleaned_clinic$days_to_death != "[Not Applicable]",
                                      cleaned_clinic$days_to_death,
                                      cleaned_clinic$days_to_last_follow_up)
```

If we check the data, we see that the values in `survival_time` are strings from the “ ” around each number. We need to convert the values to numeric in order to use them in the KM plot.

```
cleaned_clinic$survival_time <- as.numeric(cleaned_clinic$survival_time)
```

Finally, we need to make a death event column that contains T if a patient has died, and F if the patient is alive. (Hint: `cleaned_clinic$vital_status` has “Alive” if a patient is alive and “Dead” if a patient passed.)

```
cleaned_clinic$death_event <- cleaned_clinic$vital_status == "Dead"
```

2. Create Survminer Objects

Run the following code block.

```
survival_object <- Surv(time = cleaned_clinic$survival_time,
                       event = cleaned_clinic$death_event)

fit_object <- survfit(survival_object ~ age_category, data = cleaned_clinic)
```

3. Create the KM Plot

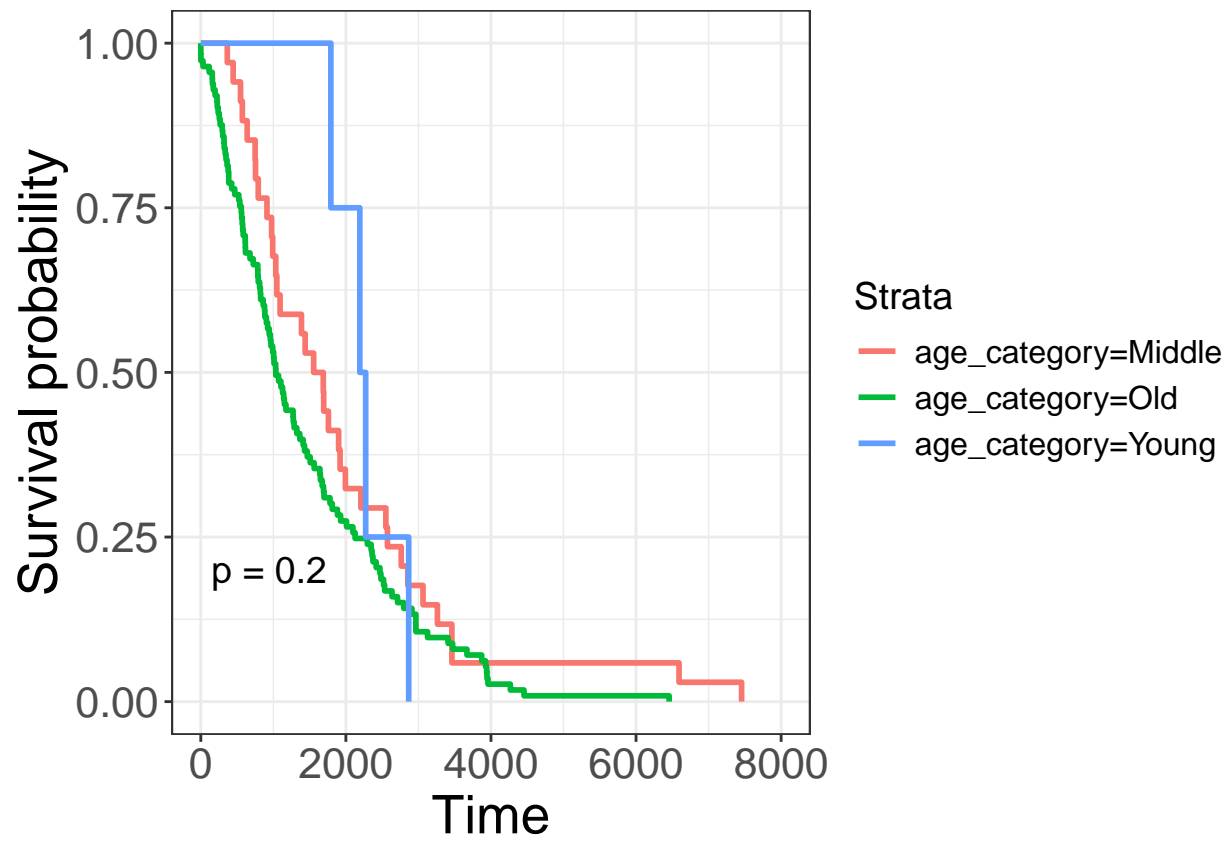
Run the following code block. The KM plot should appear in the plot tab on the right. Manually expand the tab if you get an error about the plot being too small.

```
survplot <- ggsurvplot(fit_object,
                      pval=TRUE,
                      ggtheme = theme(plot.margin = unit(c(1,1,1,1), "cm")),
                      legend = 'right')
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## i The deprecated feature was likely used in the ggpubr package.
## Please report the issue at <https://github.com/kassambara/ggpubr/issues>.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
KM_plot <- survplot$plot + theme_bw() + theme(axis.title = element_text(size=20),
                                             axis.text = element_text(size=16),
                                             legend.title = element_text(size=14),
                                             legend.text = element_text(size=12))
```

KM_plot



Answer the following questions: # 1. What does the y axis represent?

The y-axis represents the survival probability, starting at 1.0 (100% survival) and decreases over time.

2. What does the x axis represent?

The x-axis represents time in days since diagnosis. It shows how long patients were followed up, up to around 8000 days or 22 years).

3. What age group appears to have the greatest rate of survival?

The Young group appears to have the greatest survival rate with its curve staying higher than the Middle and Old groups for most of the follow-up period.

4. What age group appears to have the worst rate of survival?

The Old group shows the worst survival rate with its curve dropping the fastest, indicating higher mortality compared to the other groups.

5. What is the approx. 15-year survival rate for “Old” breast cancer patients?

The approximate 15-year survival rate for “Old” patients is between 0–5%.

6. What is the p-value? Is it statistically significant and what does that tell us about the relationship between age and breast cancer survival?

The p-value is 0.2, which is not statistically significant, meaning the relationship between age and breast cancer survival is not statistically significant, may be due to chance.

(6) Submitting Coding HW

Your homework should be a secure copy (scp) of the original file in fa24_course_materials. Push the file to Github through your qbio_490_name local repository and submit a link to your github repo to Blackboard.