

WIE2003 Alternative Assessment 2

Lim Zheng Yu

2022-06-02

(a)

Find and get a dataset from the datasets available within R. Perform exploratory data analysis (EDA) and prepare a codebook on that dataset using a newer method in R. Label your work clearly on EDA and codebook.

The dataset chosen from the datasets library is “quakes”. This data set shows the locations of 1000 seismic events (i.e. earthquakes) of Richter Magnitude of above 4.0 near Fiji since 1964.

```
## 'data.frame':    1000 obs. of  6 variables:
## $ No.          : chr  "1" "2" "3" "4" ...
## $ lat          : num  -20.4 -20.6 -26 -18 -20.4 ...
## $ long         : num  182 181 184 182 182 ...
## $ depth        : int  562 650 42 626 649 195 82 194 211 622 ...
## $ mag          : num  4.8 4.2 5.4 4.1 4 4 4.8 4.4 4.7 4.3 ...
## $ stations: int  41 15 43 19 11 12 43 15 35 19 ...
```

The first few observations of this dataset are as below:

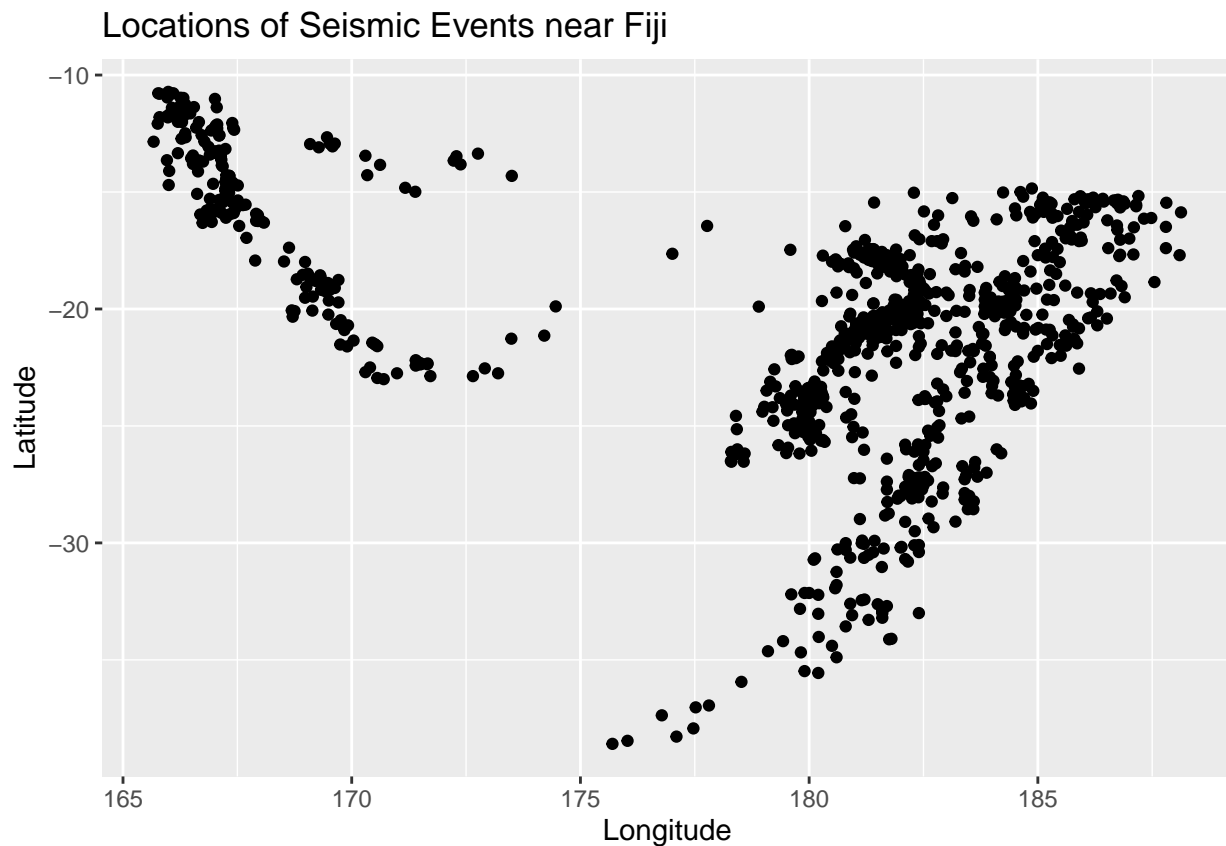
```
##   No.    lat    long depth mag stations
## 1    1 -20.42 181.62   562 4.8         41
## 2    2 -20.62 181.03   650 4.2         15
## 3    3 -26.00 184.10    42 5.4         43
## 4    4 -17.97 181.66   626 4.1         19
## 5    5 -20.42 181.96   649 4.0         11
## 6    6 -19.68 184.31   195 4.0         12
```

A summary of the dataset

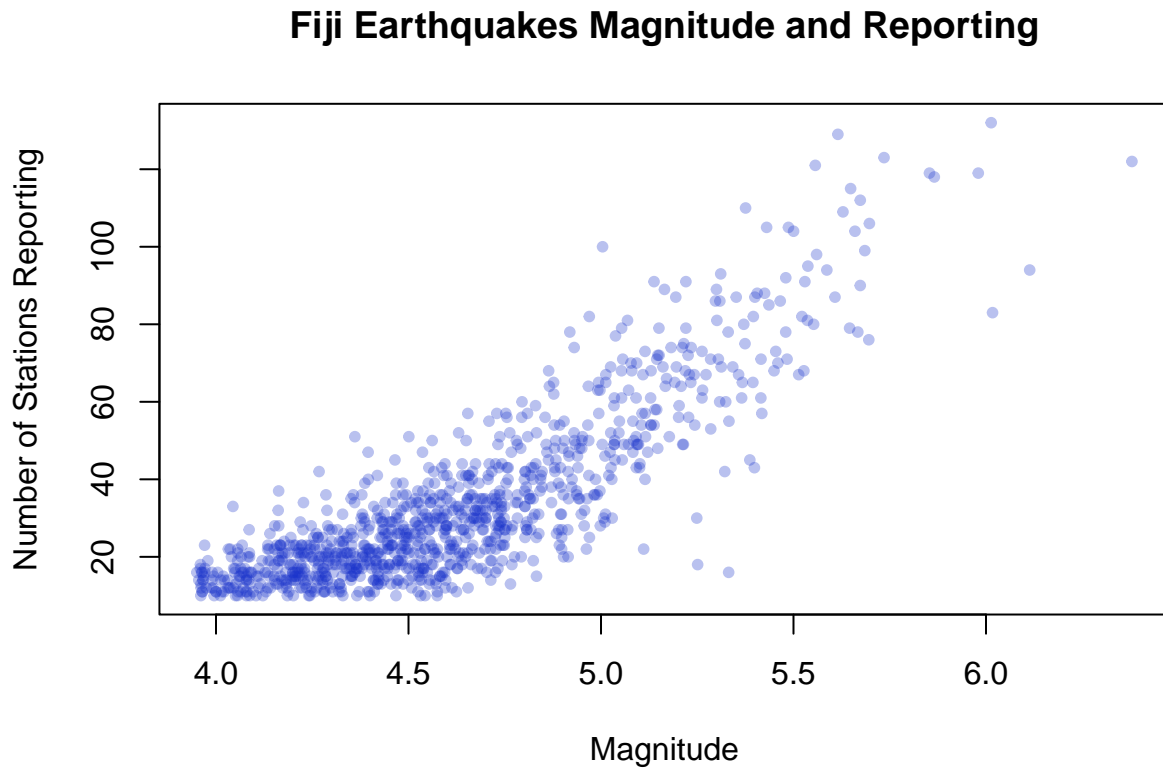
```
##      No.          lat          long          depth
## Length:1000      Min.    :-38.59   Min.    :165.7   Min.    : 40.0
## Class :character 1st Qu.: -23.47   1st Qu.:179.6   1st Qu.: 99.0
## Mode  :character Median  :-20.30   Median :181.4   Median :247.0
##                      Mean    :-20.64   Mean    :179.5   Mean    :311.4
##                      3rd Qu.: -17.64   3rd Qu.:183.2   3rd Qu.:543.0
##                      Max.     :-10.72   Max.     :188.1   Max.     :680.0
##      mag      stations
## Min.    :4.00   Min.    : 10.00
## 1st Qu.:4.30   1st Qu.: 18.00
## Median :4.60   Median : 27.00
## Mean    :4.62   Mean    : 33.42
## 3rd Qu.:4.90   3rd Qu.: 42.00
## Max.    :6.40   Max.    :132.00
```

This plot shows the approximate location of those seismic events according to their longitude and latitude

```
## package 'tidyverse' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\limzh\AppData\Local\Temp\RtmpshrVxN\downloaded_packages
```



The below figure shows the correlation between the magnitude of earthquakes and the number of stations reporting those events



The above figures show that:

- Most seismic activity occurs at locations with approximate latitude of -10 to -20, longitude of 165 to 170 and latitude of -15 to -25, longitude of 180 to 185.
- The higher the magnitude of the earthquakes, the higher number of stations are able to report those earthquakes.
- As the magnitude of seismic activity increases, the number of seismic activity decreases.

Codebook for df

Autogenerated data summary from dataMaid

2022-06-02 15:46:42

Data report overview

The dataset examined has the following dimensions:

Feature	Result
Number of observations	1000
Number of variables	6

Codebook summary table

Label	Variable	Class	# unique values	Missing	Description
	No.	character	1000	0.00 %	
	lat	numeric	721	0.00 %	
	long	numeric	605	0.00 %	
	depth	integer	422	0.00 %	
	mag	numeric	22	0.00 %	
	stations	integer	102	0.00 %	

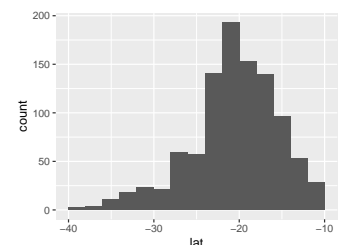
Variable list

No.

- The variable is a key (distinct values for each observation).

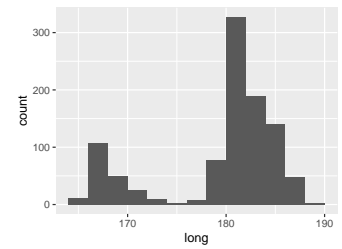
lat

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	721
Median	-20.3
1st and 3rd quartiles	-23.47; -17.64
Min. and max.	-38.59; -10.72



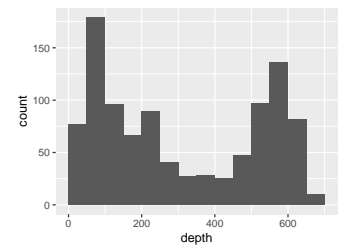
long

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	605
Median	181.41
1st and 3rd quartiles	179.62; 183.2
Min. and max.	165.67; 188.13



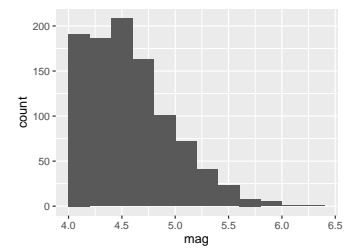
depth

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	422
Median	247
1st and 3rd quartiles	99; 543
Min. and max.	40; 680



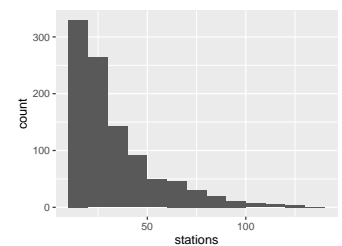
mag

Feature	Result
Variable type	numeric
Number of missing obs.	0 (0 %)
Number of unique values	22
Median	4.6
1st and 3rd quartiles	4.3; 4.9
Min. and max.	4; 6.4



stations

Feature	Result
Variable type	integer
Number of missing obs.	0 (0 %)
Number of unique values	102
Median	27
1st and 3rd quartiles	18; 42
Min. and max.	10; 132



Report generation information:

- Created by: millivan (username: limzh).
- Report creation time: Thu Jun 02 2022 15:46:43
- Report was run from directory: C:/Users/limzh/OneDrive/Desktop/UM/Y1S2/WIE2003/AAQ2
- dataMaid v1.4.1 [Pkg: 2021-10-08 from CRAN (R 4.1.3)]
- R version 4.1.3 (2022-03-10).
- Platform: x86_64-w64-mingw32/x64 (64-bit)(Windows 10 x64 (build 19044)).
- Function call: `dataMaid::makeDataReport(data = df, mode = c("summarize", "visualize", "check"), smartNum = FALSE, file = "codebook_df.Rmd", checks = list(character = "showAllFactorLevels", factor = "showAllFactorLevels", labelled = "showAllFactorLevels", haven_labelled = "showAllFactorLevels", numeric = NULL, integer = NULL, logical = NULL, Date = NULL), listChecks = FALSE, maxProbVals = Inf, codebook = TRUE, reportTitle = "Codebook for df")`

(b)

Demonstrate these FIVE (5) functions of dplyr for data manipulation:

- `filter ()`
- `arrange ()`
- `mutate ()`
- `select ()`
- `summarise ()`

You can create your own sensible dataset in certain context for this question with at least 15 observations (rows) and 4 features (columns) or you can get any suitable dataset online. Show the R code and provide a short explanation on what each function does for each produced output.

Lawyers' Ratings of State Judges in the US Superior Court

##	CONT	INTG	DMNR	DILG	CFMG	DECI	PREP	FAMI	ORAL	WRIT	PHYS	RTEN
## AARONSON, L.H.	5.7	7.9	7.7	7.3	7.1	7.4	7.1	7.1	7.1	7.0	8.3	7.8
## ALEXANDER, J.M.	6.8	8.9	8.8	8.5	7.8	8.1	8.0	8.0	7.8	7.9	8.5	8.7
## ARMENTANO, A.J.	7.2	8.1	7.8	7.8	7.5	7.6	7.5	7.5	7.3	7.4	7.9	7.8
## BERDON, R.I.	6.8	8.8	8.5	8.8	8.3	8.5	8.7	8.7	8.4	8.5	8.8	8.7
## BRACKEN, J.J.	7.3	6.4	4.3	6.5	6.0	6.2	5.7	5.7	5.1	5.3	5.5	4.8
## BURNS, E.B.	6.2	8.8	8.7	8.5	7.9	8.0	8.1	8.0	8.0	8.0	8.6	8.6
## CALLAHAN, R.J.	10.6	9.0	8.9	8.7	8.5	8.5	8.5	8.5	8.6	8.4	9.1	9.0
## COHEN, S.S.	7.0	5.9	4.9	5.1	5.4	5.9	4.8	5.1	4.7	4.9	6.8	5.0
## DALY, J.J.	7.3	8.9	8.9	8.7	8.6	8.5	8.4	8.4	8.4	8.5	8.8	8.8
## DANNEHY, J.F.	8.2	7.9	6.7	8.1	7.9	8.0	7.9	8.1	7.7	7.8	8.5	7.9
## DEAN, H.H.	7.0	8.0	7.6	7.4	7.3	7.5	7.1	7.2	7.1	7.2	8.4	7.7
## DEVITA, H.J.	6.5	8.0	7.6	7.2	7.0	7.1	6.9	7.0	7.0	7.1	6.9	7.2
## DRISCOLL, P.J.	6.7	8.6	8.2	6.8	6.9	6.6	7.1	7.3	7.2	7.2	8.1	7.7
## GRILLO, A.E.	7.0	7.5	6.4	6.8	6.5	7.0	6.6	6.8	6.3	6.6	6.2	6.5
## HADDEN, W.L. JR.	6.5	8.1	8.0	8.0	7.9	8.0	7.9	7.8	7.8	7.8	8.4	8.0
## HAMILL, E.C.	7.3	8.0	7.4	7.7	7.3	7.3	7.3	7.2	7.1	7.2	8.0	7.6
## HEALEY, A.H.	8.0	7.6	6.6	7.2	6.5	6.5	6.8	6.7	6.4	6.5	6.9	6.7
## HULL, T.C.	7.7	7.7	6.7	7.5	7.4	7.5	7.1	7.3	7.1	7.3	8.1	7.4
## LEVINE, I.	8.3	8.2	7.4	7.8	7.7	7.7	7.7	7.8	7.5	7.6	8.0	8.0
## LEVISTER, R.L.	9.6	6.9	5.7	6.6	6.9	6.6	6.2	6.0	5.8	5.8	7.2	6.0
## MARTIN, L.F.	7.1	8.2	7.7	7.1	6.6	6.6	6.7	6.7	6.8	6.8	7.5	7.3
## MCGRATH, J.F.	7.6	7.3	6.9	6.8	6.7	6.8	6.4	6.3	6.3	6.3	7.4	6.6
## MIGNONE, A.F.	6.6	7.4	6.2	6.2	5.4	5.7	5.8	5.9	5.2	5.8	4.7	5.2
## MISSAL, H.M.	6.2	8.3	8.1	7.7	7.4	7.3	7.3	7.3	7.2	7.3	7.8	7.6
## MULVEY, H.M.	7.5	8.7	8.5	8.6	8.5	8.4	8.5	8.5	8.4	8.4	8.7	8.7
## NARUK, H.J.	7.8	8.9	8.7	8.9	8.7	8.8	8.9	9.0	8.8	8.9	9.0	9.0
## O'BRIEN, F.J.	7.1	8.5	8.3	8.0	7.9	7.9	7.8	7.8	7.8	7.7	8.3	8.2
## O'SULLIVAN, T.J.	7.5	9.0	8.9	8.7	8.4	8.5	8.4	8.3	8.3	8.3	8.8	8.7
## PASKEY, L.	7.5	8.1	7.7	8.2	8.0	8.1	8.2	8.4	8.0	8.1	8.4	8.1
## RUBINOW, J.E.	7.1	9.2	9.0	9.0	8.4	8.6	9.1	9.1	8.9	9.0	8.9	9.2
## SADEN, G.A.	6.6	7.4	6.9	8.4	8.0	7.9	8.2	8.4	7.7	7.9	8.4	7.5
## SATANIELLO, A.G.	8.4	8.0	7.9	7.9	7.8	7.8	7.6	7.4	7.4	7.4	8.1	7.9
## SHEA, D.M.	6.9	8.5	7.8	8.5	8.1	8.2	8.4	8.5	8.1	8.3	8.7	8.3
## SHEA, J.F. JR.	7.3	8.9	8.8	8.7	8.4	8.5	8.5	8.5	8.4	8.4	8.8	8.8
## SIDOR, W.J.	7.7	6.2	5.1	5.6	5.6	5.9	5.6	5.6	5.3	5.5	6.3	5.3
## SPEZIALE, J.A.	8.5	8.3	8.1	8.3	8.4	8.2	8.2	8.1	7.9	8.0	8.0	8.2
## SPONZO, M.J.	6.9	8.3	8.0	8.1	7.9	7.9	7.9	7.7	7.6	7.7	8.1	8.0
## STAPLETON, J.F.	6.5	8.2	7.7	7.8	7.6	7.7	7.7	7.7	7.5	7.6	8.5	7.7
## TESTO, R.J.	8.3	7.3	7.0	6.8	7.0	7.1	6.7	6.7	6.7	6.7	8.0	7.0
## TIERNEY, W.L. JR.	8.3	8.2	7.8	8.3	8.4	8.3	7.7	7.6	7.5	7.7	8.1	7.9
## WALL, R.A.	9.0	7.0	5.9	7.0	7.0	7.2	6.9	6.9	6.5	6.6	7.6	6.6
## WRIGHT, D.B.	7.1	8.4	8.4	7.7	7.5	7.7	7.8	8.2	8.0	8.1	8.3	8.1
## ZARRILLI, K.J.	8.6	7.4	7.0	7.5	7.5	7.7	7.4	7.2	6.9	7.0	7.8	7.1

[,1] CONT Number of contacts of lawyer with judge. [,2] INTG Judicial integrity. [,3] DMNR Demeanor. [,4] DILG Diligence. [,5] CFMG Case flow managing. [,6] DECI Prompt decisions. [,7] PREP Preparation for trial. [,8] FAMI Familiarity with law. [,9] ORAL Sound oral rulings. [,10] WRIT Sound written rulings. [,11] PHYS Physical ability. [,12] RTEN Worthy of retention.

i. filter()

The filter() function allows us to choose data from the dataset which satisfy the provided conditions.

```
# When passed in the condition of DILG >= 8.0, only Judges  
# with a diligence rating of 8.0 and above are shown  
us %>% filter(DILG >= 8.0)
```

##	CONT	INTG	DMNR	DILG	CFMG	DECI	PREP	FAMI	ORAL	WRIT	PHYS	RTEN
## ALEXANDER, J.M.	6.8	8.9	8.8	8.5	7.8	8.1	8.0	8.0	7.8	7.9	8.5	8.7
## BERDON, R.I.	6.8	8.8	8.5	8.8	8.3	8.5	8.7	8.7	8.4	8.5	8.8	8.7
## BURNS, E.B.	6.2	8.8	8.7	8.5	7.9	8.0	8.1	8.0	8.0	8.0	8.6	8.6
## CALLAHAN, R.J.	10.6	9.0	8.9	8.7	8.5	8.5	8.5	8.5	8.6	8.4	9.1	9.0
## DALY, J.J.	7.3	8.9	8.9	8.7	8.6	8.5	8.4	8.4	8.4	8.5	8.8	8.8
## DANNEHY, J.F.	8.2	7.9	6.7	8.1	7.9	8.0	7.9	8.1	7.7	7.8	8.5	7.9
## HADDEN, W.L. JR.	6.5	8.1	8.0	8.0	7.9	8.0	7.9	7.8	7.8	7.8	8.4	8.0
## MULVEY, H.M.	7.5	8.7	8.5	8.6	8.5	8.4	8.5	8.5	8.4	8.4	8.7	8.7
## NARUK, H.J.	7.8	8.9	8.7	8.9	8.7	8.8	8.9	9.0	8.8	8.9	9.0	9.0
## O'BRIEN, F.J.	7.1	8.5	8.3	8.0	7.9	7.9	7.8	7.8	7.8	7.7	8.3	8.2
## O'SULLIVAN, T.J.	7.5	9.0	8.9	8.7	8.4	8.5	8.4	8.3	8.3	8.3	8.8	8.7
## PASKEY, L.	7.5	8.1	7.7	8.2	8.0	8.1	8.2	8.4	8.0	8.1	8.4	8.1
## RUBINOW, J.E.	7.1	9.2	9.0	9.0	8.4	8.6	9.1	9.1	8.9	9.0	8.9	9.2
## SADEN, G.A.	6.6	7.4	6.9	8.4	8.0	7.9	8.2	8.4	7.7	7.9	8.4	7.5
## SHEA, D.M.	6.9	8.5	7.8	8.5	8.1	8.2	8.4	8.5	8.1	8.3	8.7	8.3
## SHEA, J.F. JR.	7.3	8.9	8.8	8.7	8.4	8.5	8.5	8.5	8.4	8.4	8.8	8.8
## SPEZIALE, J.A.	8.5	8.3	8.1	8.3	8.4	8.2	8.2	8.1	7.9	8.0	8.0	8.2
## SPONZO, M.J.	6.9	8.3	8.0	8.1	7.9	7.9	7.9	7.7	7.6	7.7	8.1	8.0
## TIERNEY, W.L. JR.	8.3	8.2	7.8	8.3	8.4	8.3	7.7	7.6	7.5	7.7	8.1	7.9

```
# we can separate multiple conditions using a comma which functions as an AND operator  
us %>% filter(PHYS <= 7.5, RTEN >= 6.5)
```

##	CONT	INTG	DMNR	DILG	CFMG	DECI	PREP	FAMI	ORAL	WRIT	PHYS	RTEN
## DEVITA, H.J.	6.5	8.0	7.6	7.2	7.0	7.1	6.9	7.0	7.0	7.1	6.9	7.2
## GRILLO, A.E.	7.0	7.5	6.4	6.8	6.5	7.0	6.6	6.8	6.3	6.6	6.2	6.5
## HEALEY, A.H.	8.0	7.6	6.6	7.2	6.5	6.5	6.8	6.7	6.4	6.5	6.9	6.7
## MARTIN, L.F.	7.1	8.2	7.7	7.1	6.6	6.6	6.7	6.7	6.8	6.8	7.5	7.3
## MCGRATH, J.F.	7.6	7.3	6.9	6.8	6.7	6.8	6.4	6.3	6.3	6.3	7.4	6.6

ii. arrange()

The arrange() function takes in column names as parameters and orders the data accordingly.

```
# CONT attribute is passed in as a parameter.  
# Hence the data will arrange according to the value of CONT.  
us %>% arrange(CONT)
```

##	CONT	INTG	DMNR	DILG	CFMG	DECI	PREP	FAMI	ORAL	WRIT	PHYS	RTEN
## AARONSON, L.H.	5.7	7.9	7.7	7.3	7.1	7.4	7.1	7.1	7.1	7.0	8.3	7.8
## BURNS, E.B.	6.2	8.8	8.7	8.5	7.9	8.0	8.1	8.0	8.0	8.0	8.6	8.6
## MISSAL, H.M.	6.2	8.3	8.1	7.7	7.4	7.3	7.3	7.3	7.2	7.3	7.8	7.6
## DEVITA, H.J.	6.5	8.0	7.6	7.2	7.0	7.1	6.9	7.0	7.0	7.1	6.9	7.2
## HADDEN, W.L. JR.	6.5	8.1	8.0	8.0	7.9	8.0	7.9	7.8	7.8	7.8	8.4	8.0
## STAPLETON, J.F.	6.5	8.2	7.7	7.8	7.6	7.7	7.7	7.7	7.5	7.6	8.5	7.7
## MIGNONE, A.F.	6.6	7.4	6.2	6.2	5.4	5.7	5.8	5.9	5.2	5.8	4.7	5.2
## SADEN, G.A.	6.6	7.4	6.9	8.4	8.0	7.9	8.2	8.4	7.7	7.9	8.4	7.5
## DRISCOLL, P.J.	6.7	8.6	8.2	6.8	6.9	6.6	7.1	7.3	7.2	7.2	8.1	7.7
## ALEXANDER, J.M.	6.8	8.9	8.8	8.5	7.8	8.1	8.0	8.0	7.8	7.9	8.5	8.7
## BERDON, R.I.	6.8	8.8	8.5	8.8	8.3	8.5	8.7	8.7	8.4	8.5	8.8	8.7
## SHEA, D.M.	6.9	8.5	7.8	8.5	8.1	8.2	8.4	8.5	8.1	8.3	8.7	8.3
## SPONZO, M.J.	6.9	8.3	8.0	8.1	7.9	7.9	7.9	7.7	7.6	7.7	8.1	8.0
## COHEN, S.S.	7.0	5.9	4.9	5.1	5.4	5.9	4.8	5.1	4.7	4.9	6.8	5.0
## DEAN, H.H.	7.0	8.0	7.6	7.4	7.3	7.5	7.1	7.2	7.1	7.2	8.4	7.7
## GRILLO, A.E.	7.0	7.5	6.4	6.8	6.5	7.0	6.6	6.8	6.3	6.6	6.2	6.5
## MARTIN, L.F.	7.1	8.2	7.7	7.1	6.6	6.6	6.7	6.7	6.8	6.8	7.5	7.3
## O'BRIEN, F.J.	7.1	8.5	8.3	8.0	7.9	7.9	7.8	7.8	7.8	7.7	8.3	8.2
## RUBINOW, J.E.	7.1	9.2	9.0	9.0	8.4	8.6	9.1	9.1	8.9	9.0	8.9	9.2
## WRIGHT, D.B.	7.1	8.4	8.4	7.7	7.5	7.7	7.8	8.2	8.0	8.1	8.3	8.1
## ARMENTANO, A.J.	7.2	8.1	7.8	7.8	7.5	7.6	7.5	7.5	7.3	7.4	7.9	7.8
## BRACKEN, J.J.	7.3	6.4	4.3	6.5	6.0	6.2	5.7	5.7	5.1	5.3	5.5	4.8
## DALY, J.J.	7.3	8.9	8.9	8.7	8.6	8.5	8.4	8.4	8.4	8.5	8.8	8.8
## HAMILL, E.C.	7.3	8.0	7.4	7.7	7.3	7.3	7.3	7.2	7.1	7.2	8.0	7.6
## SHEA, J.F. JR.	7.3	8.9	8.8	8.7	8.4	8.5	8.5	8.5	8.4	8.4	8.8	8.8
## MULVEY, H.M.	7.5	8.7	8.5	8.6	8.5	8.4	8.5	8.5	8.4	8.4	8.7	8.7
## O'SULLIVAN, T.J.	7.5	9.0	8.9	8.7	8.4	8.5	8.4	8.3	8.3	8.3	8.8	8.7
## PASKEY, L.	7.5	8.1	7.7	8.2	8.0	8.1	8.2	8.4	8.0	8.1	8.4	8.1
## MCGRATH, J.F.	7.6	7.3	6.9	6.8	6.7	6.8	6.4	6.3	6.3	6.3	7.4	6.6
## HULL, T.C.	7.7	7.7	6.7	7.5	7.4	7.5	7.1	7.3	7.1	7.3	8.1	7.4
## SIDOR, W.J.	7.7	6.2	5.1	5.6	5.6	5.9	5.6	5.6	5.3	5.5	6.3	5.3
## NARUK, H.J.	7.8	8.9	8.7	8.9	8.7	8.8	8.9	9.0	8.8	8.9	9.0	9.0
## HEALEY, A.H.	8.0	7.6	6.6	7.2	6.5	6.5	6.8	6.7	6.4	6.5	6.9	6.7
## DANNEHY, J.F.	8.2	7.9	6.7	8.1	7.9	8.0	7.9	8.1	7.7	7.8	8.5	7.9
## LEVINE, I.	8.3	8.2	7.4	7.8	7.7	7.7	7.7	7.8	7.5	7.6	8.0	8.0
## TESTO, R.J.	8.3	7.3	7.0	6.8	7.0	7.1	6.7	6.7	6.7	6.7	8.0	7.0
## TIERNEY, W.L. JR.	8.3	8.2	7.8	8.3	8.4	8.3	7.7	7.6	7.5	7.7	8.1	7.9
## SATANIELLO, A.G.	8.4	8.0	7.9	7.9	7.8	7.8	7.6	7.4	7.4	7.4	8.1	7.9
## SPEZIALE, J.A.	8.5	8.3	8.1	8.3	8.4	8.2	8.2	8.1	7.9	8.0	8.0	8.2
## ZARRILLI, K.J.	8.6	7.4	7.0	7.5	7.5	7.7	7.4	7.2	6.9	7.0	7.8	7.1
## WALL, R.A.	9.0	7.0	5.9	7.0	7.0	7.2	6.9	6.9	6.5	6.6	7.6	6.6
## LEVISTER, R.L.	9.6	6.9	5.7	6.6	6.9	6.6	6.2	6.0	5.8	5.8	7.2	6.0
## CALLAHAN, R.J.	10.6	9.0	8.9	8.7	8.5	8.5	8.5	8.5	8.6	8.4	9.1	9.0

When multiple arguments are provided, the additional parameter will be used
to break ties of the values of preceding parameters.
us %>% arrange(INTG, DMNR)

##	CONT	INTG	DMNR	DILG	CFMG	DECI	PREP	FAMI	ORAL	WRIT	PHYS	RTEN
## COHEN,S.S.	7.0	5.9	4.9	5.1	5.4	5.9	4.8	5.1	4.7	4.9	6.8	5.0
## SIDOR,W.J.	7.7	6.2	5.1	5.6	5.6	5.9	5.6	5.6	5.3	5.5	6.3	5.3
## BRACKEN,J.J.	7.3	6.4	4.3	6.5	6.0	6.2	5.7	5.7	5.1	5.3	5.5	4.8
## LEVISTER,R.L.	9.6	6.9	5.7	6.6	6.9	6.6	6.2	6.0	5.8	5.8	7.2	6.0
## WALL,R.A.	9.0	7.0	5.9	7.0	7.0	7.2	6.9	6.9	6.5	6.6	7.6	6.6
## MCGRATH,J.F.	7.6	7.3	6.9	6.8	6.7	6.8	6.4	6.3	6.3	6.3	7.4	6.6
## TESTO,R.J.	8.3	7.3	7.0	6.8	7.0	7.1	6.7	6.7	6.7	6.7	8.0	7.0
## MIGNONE,A.F.	6.6	7.4	6.2	6.2	5.4	5.7	5.8	5.9	5.2	5.8	4.7	5.2
## SADEN,G.A.	6.6	7.4	6.9	8.4	8.0	7.9	8.2	8.4	7.7	7.9	8.4	7.5
## ZARRILLI,K.J.	8.6	7.4	7.0	7.5	7.5	7.7	7.4	7.2	6.9	7.0	7.8	7.1
## GRILLO,A.E.	7.0	7.5	6.4	6.8	6.5	7.0	6.6	6.8	6.3	6.6	6.2	6.5
## HEALEY,A.H.	8.0	7.6	6.6	7.2	6.5	6.5	6.8	6.7	6.4	6.5	6.9	6.7
## HULL,T.C.	7.7	7.7	6.7	7.5	7.4	7.5	7.1	7.3	7.1	7.3	8.1	7.4
## DANNEHY,J.F.	8.2	7.9	6.7	8.1	7.9	8.0	7.9	8.1	7.7	7.8	8.5	7.9
## AARONSON,L.H.	5.7	7.9	7.7	7.3	7.1	7.4	7.1	7.1	7.1	7.0	8.3	7.8
## HAMILL,E.C.	7.3	8.0	7.4	7.7	7.3	7.3	7.3	7.2	7.1	7.2	8.0	7.6
## DEAN,H.H.	7.0	8.0	7.6	7.4	7.3	7.5	7.1	7.2	7.1	7.2	8.4	7.7
## DEVITA,H.J.	6.5	8.0	7.6	7.2	7.0	7.1	6.9	7.0	7.0	7.1	6.9	7.2
## SATANIELLO,A.G.	8.4	8.0	7.9	7.9	7.8	7.8	7.6	7.4	7.4	7.4	8.1	7.9
## PASKEY,L.	7.5	8.1	7.7	8.2	8.0	8.1	8.2	8.4	8.0	8.1	8.4	8.1
## ARMENTANO,A.J.	7.2	8.1	7.8	7.8	7.5	7.6	7.5	7.5	7.3	7.4	7.9	7.8
## HADDEN,W.L.JR.	6.5	8.1	8.0	8.0	7.9	8.0	7.9	7.8	7.8	7.8	8.4	8.0
## LEVINE,I.	8.3	8.2	7.4	7.8	7.7	7.7	7.7	7.8	7.5	7.6	8.0	8.0
## MARTIN,L.F.	7.1	8.2	7.7	7.1	6.6	6.6	6.7	6.7	6.8	6.8	7.5	7.3
## STAPLETON,J.F.	6.5	8.2	7.7	7.8	7.6	7.7	7.7	7.7	7.5	7.6	8.5	7.7
## TIERNEY,W.L.JR.	8.3	8.2	7.8	8.3	8.4	8.3	7.7	7.6	7.5	7.7	8.1	7.9
## SPONZO,M.J.	6.9	8.3	8.0	8.1	7.9	7.9	7.9	7.7	7.6	7.7	8.1	8.0
## MISSAL,H.M.	6.2	8.3	8.1	7.7	7.4	7.3	7.3	7.3	7.2	7.3	7.8	7.6
## SPEZIALE,J.A.	8.5	8.3	8.1	8.3	8.4	8.2	8.2	8.1	7.9	8.0	8.0	8.2
## WRIGHT,D.B.	7.1	8.4	8.4	7.7	7.5	7.7	7.8	8.2	8.0	8.1	8.3	8.1
## SHEA,D.M.	6.9	8.5	7.8	8.5	8.1	8.2	8.4	8.5	8.1	8.3	8.7	8.3
## O'BRIEN,F.J.	7.1	8.5	8.3	8.0	7.9	7.9	7.8	7.8	7.8	7.7	8.3	8.2
## DRISCOLL,P.J.	6.7	8.6	8.2	6.8	6.9	6.6	7.1	7.3	7.2	7.2	8.1	7.7
## MULVEY,H.M.	7.5	8.7	8.5	8.6	8.5	8.4	8.5	8.5	8.4	8.4	8.7	8.7
## BERDON,R.I.	6.8	8.8	8.5	8.8	8.3	8.5	8.7	8.7	8.4	8.5	8.8	8.7
## BURNS,E.B.	6.2	8.8	8.7	8.5	7.9	8.0	8.1	8.0	8.0	8.0	8.6	8.6
## NARUK,H.J.	7.8	8.9	8.7	8.9	8.7	8.8	8.9	9.0	8.8	8.9	9.0	9.0
## ALEXANDER,J.M.	6.8	8.9	8.8	8.5	7.8	8.1	8.0	8.0	7.8	7.9	8.5	8.7
## SHEA,J.F.JR.	7.3	8.9	8.8	8.7	8.4	8.5	8.5	8.5	8.4	8.4	8.8	8.8
## DALY,J.J.	7.3	8.9	8.9	8.7	8.6	8.5	8.4	8.4	8.4	8.5	8.8	8.8
## CALLAHAN,R.J.	10.6	9.0	8.9	8.7	8.5	8.5	8.5	8.5	8.6	8.4	9.1	9.0
## O'SULLIVAN,T.J.	7.5	9.0	8.9	8.7	8.4	8.5	8.4	8.3	8.3	8.3	8.8	8.7
## RUBINOW,J.E.	7.1	9.2	9.0	9.0	8.4	8.6	9.1	9.1	8.9	9.0	8.9	9.2

iii. mutate()

The mutate() function adds a new column to the dataset which is based on existing column from the dataset. The provided parameter will be the name of the new column and the value of the parameter is the definition of it.

```
us %>% mutate(DEC1_PCTG = (DEC1 / 10) * 100)
```

##	CONT	INTG	DMNR	DILG	CFMG	DEC1	PREP	FAMI	ORAL	WRIT	PHYS	RTEN
## AARONSON, L.H.	5.7	7.9	7.7	7.3	7.1	7.4	7.1	7.1	7.1	7.0	8.3	7.8
## ALEXANDER, J.M.	6.8	8.9	8.8	8.5	7.8	8.1	8.0	8.0	7.8	7.9	8.5	8.7
## ARMENTANO, A.J.	7.2	8.1	7.8	7.8	7.5	7.6	7.5	7.5	7.3	7.4	7.9	7.8
## BERDON, R.I.	6.8	8.8	8.5	8.8	8.3	8.5	8.7	8.7	8.4	8.5	8.8	8.7
## BRACKEN, J.J.	7.3	6.4	4.3	6.5	6.0	6.2	5.7	5.7	5.1	5.3	5.5	4.8
## BURNS, E.B.	6.2	8.8	8.7	8.5	7.9	8.0	8.1	8.0	8.0	8.0	8.6	8.6
## CALLAHAN, R.J.	10.6	9.0	8.9	8.7	8.5	8.5	8.5	8.5	8.6	8.4	9.1	9.0
## COHEN, S.S.	7.0	5.9	4.9	5.1	5.4	5.9	4.8	5.1	4.7	4.9	6.8	5.0
## DALY, J.J.	7.3	8.9	8.9	8.7	8.6	8.5	8.4	8.4	8.4	8.5	8.8	8.8
## DANNEHY, J.F.	8.2	7.9	6.7	8.1	7.9	8.0	7.9	8.1	7.7	7.8	8.5	7.9
## DEAN, H.H.	7.0	8.0	7.6	7.4	7.3	7.5	7.1	7.2	7.1	7.2	8.4	7.7
## DEVITA, H.J.	6.5	8.0	7.6	7.2	7.0	7.1	6.9	7.0	7.0	7.1	6.9	7.2
## DRISCOLL, P.J.	6.7	8.6	8.2	6.8	6.9	6.6	7.1	7.3	7.2	7.2	8.1	7.7
## GRILLO, A.E.	7.0	7.5	6.4	6.8	6.5	7.0	6.6	6.8	6.3	6.6	6.2	6.5
## HADDEN, W.L. JR.	6.5	8.1	8.0	8.0	7.9	8.0	7.9	7.8	7.8	7.8	8.4	8.0
## HAMILL, E.C.	7.3	8.0	7.4	7.7	7.3	7.3	7.3	7.2	7.1	7.2	8.0	7.6
## HEALEY, A.H.	8.0	7.6	6.6	7.2	6.5	6.5	6.8	6.7	6.4	6.5	6.9	6.7
## HULL, T.C.	7.7	7.7	6.7	7.5	7.4	7.5	7.1	7.3	7.1	7.3	8.1	7.4
## LEVINE, I.	8.3	8.2	7.4	7.8	7.7	7.7	7.7	7.8	7.5	7.6	8.0	8.0
## LEVISTER, R.L.	9.6	6.9	5.7	6.6	6.9	6.6	6.2	6.0	5.8	5.8	7.2	6.0
## MARTIN, L.F.	7.1	8.2	7.7	7.1	6.6	6.6	6.7	6.7	6.8	6.8	7.5	7.3
## MCGRATH, J.F.	7.6	7.3	6.9	6.8	6.7	6.8	6.4	6.3	6.3	6.3	7.4	6.6
## MIGNONE, A.F.	6.6	7.4	6.2	6.2	5.4	5.7	5.8	5.9	5.2	5.8	4.7	5.2
## MISSAL, H.M.	6.2	8.3	8.1	7.7	7.4	7.3	7.3	7.3	7.2	7.3	7.8	7.6
## MULVEY, H.M.	7.5	8.7	8.5	8.6	8.5	8.4	8.5	8.5	8.4	8.4	8.7	8.7
## NARUK, H.J.	7.8	8.9	8.7	8.9	8.7	8.8	8.9	9.0	8.8	8.9	9.0	9.0
## O'BRIEN, F.J.	7.1	8.5	8.3	8.0	7.9	7.9	7.8	7.8	7.8	7.7	8.3	8.2
## O'SULLIVAN, T.J.	7.5	9.0	8.9	8.7	8.4	8.5	8.4	8.3	8.3	8.3	8.8	8.7
## PASKEY, L.	7.5	8.1	7.7	8.2	8.0	8.1	8.2	8.4	8.0	8.1	8.4	8.1
## RUBINOW, J.E.	7.1	9.2	9.0	9.0	8.4	8.6	9.1	9.1	8.9	9.0	8.9	9.2
## SADEN, G.A.	6.6	7.4	6.9	8.4	8.0	7.9	8.2	8.4	7.7	7.9	8.4	7.5
## SATANIELLO, A.G.	8.4	8.0	7.9	7.9	7.8	7.8	7.6	7.4	7.4	7.4	8.1	7.9
## SHEA, D.M.	6.9	8.5	7.8	8.5	8.1	8.2	8.4	8.5	8.1	8.3	8.7	8.3
## SHEA, J.F. JR.	7.3	8.9	8.8	8.7	8.4	8.5	8.5	8.5	8.4	8.4	8.8	8.8
## SIDOR, W.J.	7.7	6.2	5.1	5.6	5.6	5.9	5.6	5.6	5.3	5.5	6.3	5.3
## SPEZIALE, J.A.	8.5	8.3	8.1	8.3	8.4	8.2	8.2	8.1	7.9	8.0	8.0	8.2
## SPONZO, M.J.	6.9	8.3	8.0	8.1	7.9	7.9	7.9	7.7	7.6	7.7	8.1	8.0
## STAPLETON, J.F.	6.5	8.2	7.7	7.8	7.6	7.7	7.7	7.7	7.5	7.6	8.5	7.7
## TESTO, R.J.	8.3	7.3	7.0	6.8	7.0	7.1	6.7	6.7	6.7	6.7	8.0	7.0
## TIERNEY, W.L. JR.	8.3	8.2	7.8	8.3	8.4	8.3	7.7	7.6	7.5	7.7	8.1	7.9
## WALL, R.A.	9.0	7.0	5.9	7.0	7.0	7.2	6.9	6.9	6.5	6.6	7.6	6.6
## WRIGHT, D.B.	7.1	8.4	8.4	7.7	7.5	7.7	7.8	8.2	8.0	8.1	8.3	8.1
## ZARRILLI, K.J.	8.6	7.4	7.0	7.5	7.5	7.7	7.4	7.2	6.9	7.0	7.8	7.1

##	DECI_PCTG
## AARONSON, L. H.	74
## ALEXANDER, J. M.	81
## ARMENTANO, A. J.	76
## BERDON, R. I.	85
## BRACKEN, J. J.	62
## BURNS, E. B.	80
## CALLAHAN, R. J.	85
## COHEN, S. S.	59
## DALY, J. J.	85
## DANNEHY, J. F.	80
## DEAN, H. H.	75
## DEVITA, H. J.	71
## DRISCOLL, P. J.	66
## GRILLO, A. E.	70
## HADDEN, W. L. JR.	80
## HAMILL, E. C.	73
## HEALEY, A. H.	65
## HULL, T. C.	75
## LEVINE, I.	77
## LEVISTER, R. L.	66
## MARTIN, L. F.	66
## MCGRATH, J. F.	68
## MIGNONE, A. F.	57
## MISSAL, H. M.	73
## MULVEY, H. M.	84
## NARUK, H. J.	88
## O'BRIEN, F. J.	79
## O'SULLIVAN, T. J.	85
## PASKEY, L.	81
## RUBINOW, J. E.	86
## SADEN, G. A.	79
## SATANIELLO, A. G.	78
## SHEA, D. M.	82
## SHEA, J. F. JR.	85
## SIDOR, W. J.	59
## SPEZIALE, J. A.	82
## SPONZO, M. J.	79
## STAPLETON, J. F.	77
## TESTO, R. J.	71
## TIERNEY, W. L. JR.	83
## WALL, R. A.	72
## WRIGHT, D. B.	77
## ZARRILLI, K. J.	77

iv. select()

The select() function chooses the columns that we have passed in as parameters to be outputted.

```
us %>% mutate(DEC1_PCTG = (DEC1 / 10) * 100) %>% select(DEC1_PCTG, DEC1)
```

##	DEC1_PCTG	DEC1
## AARONSON, L.H.	74	7.4
## ALEXANDER, J.M.	81	8.1
## ARMENTANO, A.J.	76	7.6
## BERDON, R.I.	85	8.5
## BRACKEN, J.J.	62	6.2
## BURNS, E.B.	80	8.0
## CALLAHAN, R.J.	85	8.5
## COHEN, S.S.	59	5.9
## DALY, J.J.	85	8.5
## DANNEHY, J.F.	80	8.0
## DEAN, H.H.	75	7.5
## DEVITA, H.J.	71	7.1
## DRISCOLL, P.J.	66	6.6
## GRILLO, A.E.	70	7.0
## HADDEN, W.L. JR.	80	8.0
## HAMILL, E.C.	73	7.3
## HEALEY, A.H.	65	6.5
## HULL, T.C.	75	7.5
## LEVINE, I.	77	7.7
## LEVISTER, R.L.	66	6.6
## MARTIN, L.F.	66	6.6
## MCGRATH, J.F.	68	6.8
## MIGNONE, A.F.	57	5.7
## MISSAL, H.M.	73	7.3
## MULVEY, H.M.	84	8.4
## NARUK, H.J.	88	8.8
## O'BRIEN, F.J.	79	7.9
## O'SULLIVAN, T.J.	85	8.5
## PASKEY, L.	81	8.1
## RUBINOW, J.E.	86	8.6
## SADEN, G.A.	79	7.9
## SATANIELLO, A.G.	78	7.8
## SHEA, D.M.	82	8.2
## SHEA, J.F. JR.	85	8.5
## SIDOR, W.J.	59	5.9
## SPEZIALE, J.A.	82	8.2
## SPONZO, M.J.	79	7.9
## STAPLETON, J.F.	77	7.7
## TESTO, R.J.	71	7.1
## TIERNEY, W.L. JR.	83	8.3
## WALL, R.A.	72	7.2
## WRIGHT, D.B.	77	7.7
## ZARRILLI, K.J.	77	7.7

```
v. summarise()
```

The summarise() function summarizes the whole dataset into one row based on the parameters provided.

```
# The parameters provided are individual of each other  
# as it only provides the summary of its own column  
us %>% summarise(FAMI=mean(FAMI, na.rm=TRUE), ORAL=max(ORAL, na.rm=TRUE))
```

```
##           FAMI ORAL  
## 1 7.488372  8.9
```