Coding Assignment 3 - 241180618

```
getwd()
```

```
## [1] "C:/Users/ushad/Documents/GitHub/BSE658/Module 3"
```

Rdata files contain multiple datasets, while a csv contains only 1 dataset

```
library(tibble)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v purrr     1.0.2
## v forcats   1.0.0     v readr     2.1.5
## v ggplot2   3.5.1     v stringr   1.5.1
## v lubridate 1.9.3     v tidyr     1.3.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lsr)

#library(dplyr)
tibb<-read_csv("C:/Users/ushad/Downloads/Arrests.csv")
```

```
## Rows: 5226 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (5): released, colour, sex, employed, citizen
## dbl (4): rownames, year, age, checks
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
tibb
```

```
## # A tibble: 5,226 x 9
##    rownames released colour  year   age sex    employed citizen checks
##       <dbl> <chr>    <chr>  <dbl> <dbl> <chr>  <chr>    <chr>    <dbl>
##  1        1 Yes      White   2002    21 Male   Yes      Yes          3
##  2        2 No       Black   1999    17 Male   Yes      Yes          3
##  3        3 Yes      White   2000    24 Male   Yes      Yes          3
##  4        4 No       Black   2000    46 Male   Yes      Yes          1
##  5        5 Yes      Black   1999    27 Female Yes      Yes          1
```

```
##  6        6 Yes      Black   1998    16 Female Yes      Yes          0
##  7        7 Yes      White   1999    40 Male   No       Yes          0
##  8        8 Yes      White   1998    34 Female Yes      Yes          1
##  9        9 Yes      Black   2000    23 Male   Yes      Yes          4
## 10       10 Yes      White   2001    30 Male   Yes      Yes          3
## # i 5,216 more rows
```

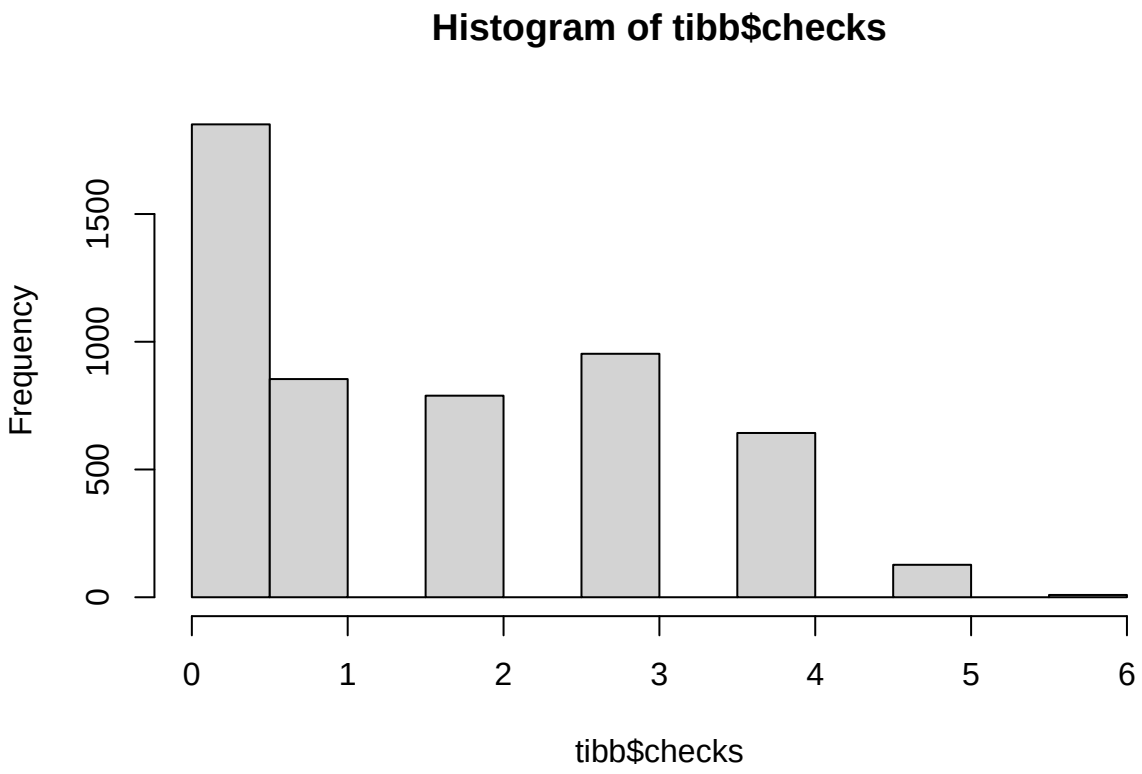
```
print(tibb$checks)
```

```
##    [1] 3 3 3 1 1 0 0 1 4 3 0 3 1 0 2 3 4 5 3 0 0 2 0 3 0 3 3 0 0 3 0 3 2 3 0 0 0
##   [38] 1 4 0 2 1 1 0 0 1 0 4 0 4 3 3 4 0 3 0 3 3 0 1 1 3 4 2 3 0 1 0 4 2 0 1 1 0
##   [75] 3 0 5 3 0 0 0 4 2 5 2 0 3 2 0 4 2 4 3 0 4 0 3 0 0 3 0 2 2 0 0 0 1 2 3 2 4
##  [112] 0 0 1 3 2 0 0 3 0 3 0 3 0 4 0 1 0 3 1 4 3 3 0 0 3 4 1 1 0 0 2 2 1 1 2 0 2
##  [149] 1 3 2 0 0 1 0 4 4 1 0 0 0 0 3 4 2 2 2 3 3 3 4 4 0 3 2 0 4 3 4 1 1 5 0 1 0
##  [186] 1 0 1 1 0 0 0 1 5 3 0 0 4 2 1 2 3 2 0 0 2 4 0 1 0 0 0 0 5 1 3 4 0 3 2 3 0
##  [223] 1 5 1 3 0 0 1 0 4 2 3 0 3 4 0 3 2 0 1 0 2 3 0 1 4 0 3 0 0 3 2 0 0 0 3 0 2
##  [260] 0 2 3 0 3 0 0 1 3 1 2 0 3 1 0 3 3 0 0 3 0 0 4 0 0 2 2 1 0 3 3 3 4 2 0 0 1
##  [297] 0 3 3 0 2 0 0 0 4 0 0 3 3 2 4 0 4 2 2 1 0 4 2 0 0 4 0 0 0 4 1 4 2 0 3 4 1
##  [334] 2 2 3 1 1 1 1 2 2 2 2 0 4 1 1 1 4 0 1 3 0 2 0 1 0 4 3 1 0 4 1 0 1 2 0 3 3
##  [371] 0 4 4 0 0 0 1 3 1 3 3 0 0 1 0 5 4 0 4 0 1 0 4 3 1 2 1 0 0 4 0 2 3 2 2 1 0
##  [408] 0 4 2 4 4 0 4 4 5 1 2 3 4 0 2 1 0 4 0 0 0 0 1 3 3 0 4 0 0 0 1 3 0 1 3 3 0
##  [445] 3 2 0 4 1 0 4 2 2 0 1 3 3 1 2 4 0 0 0 0 0 1 0 0 2 0 1 0 0 3 2 0 1 1 0 2 2
##  [482] 1 0 0 4 2 5 2 0 0 1 0 0 4 0 3 4 4 2 2 0 4 2 2 0 2 0 2 0 4 0 3 0 1 3 0 0 3
##  [519] 4 3 3 3 1 4 4 3 0 3 1 3 1 1 1 2 0 0 4 2 1 0 1 0 4 1 0 1 0 0 1 3 3 0 0 4 0
##  [556] 0 0 0 3 2 4 3 3 0 4 1 4 4 3 3 3 0 1 2 2 1 1 4 0 2 0 0 0 3 2 0 3 3 0 1 2 2
##  [593] 0 0 1 3 3 0 0 0 1 0 3 0 0 4 1 0 0 1 0 1 3 0 0 4 2 1 4 0 4 0 0 0 4 2 1 3 1
##  [630] 5 3 1 1 3 2 4 0 0 4 0 1 0 0 0 3 0 5 0 3 0 3 2 3 4 0 0 0 0 1 1 3 3 2 0 3 2
##  [667] 1 1 5 0 1 0 4 4 0 0 3 0 4 2 0 3 2 1 4 1 0 2 0 5 2 0 1 1 2 3 1 4 4 4 5 2 0
##  [704] 0 2 0 2 1 4 4 1 2 0 2 4 3 1 0 2 0 0 1 0 0 0 0 3 1 0 0 1 3 0 2 0 0 0 3 1 1
##  [741] 4 0 4 3 0 0 4 4 0 1 2 2 3 4 4 3 4 2 0 1 2 1 4 3 4 2 0 0 3 0 0 1 0 2 0 1 0
##  [778] 0 1 2 5 3 3 4 1 0 0 0 0 1 0 3 2 3 2 2 2 4 1 0 4 1 3 0 0 1 0 3 3 0 0 1 0 3
##  [815] 2 0 2 0 0 3 1 3 2 3 0 0 0 0 0 3 3 1 0 1 1 1 0 2 1 3 1 1 0 4 0 1 2 0 0 0 1
##  [852] 0 4 0 3 4 2 4 4 4 4 4 0 3 0 0 3 1 1 1 3 1 0 4 0 3 0 2 5 4 2 2 2 2 2 3 0 1 0
##  [889] 3 3 4 3 0 0 1 0 0 2 2 4 0 0 3 0 3 1 1 0 4 0 0 2 2 4 4 3 4 2 3 1 0 3 2 3 3
##  [926] 2 0 1 3 0 4 4 4 4 0 3 0 3 4 3 2 2 1 2 1 0 1 0 4 0 0 1 4 0 3 0 1 2 3 5 2 3 3
##  [963] 3 0 0 3 0 4 2 3 0 0 0 3 1 3 3 2 1 1 4 0 0 3 1 3 0 0 1 1 2 0 4 0 2 0 3 3 0
## [1000] 3 5 3 2 2 0 2 0 2 0 0 0 3 0 4 2 0 4 0 1 4 2 3 3 0 1 0 2 3 1 0 2 0 0 3 0 0
## [1037] 2 1 1 5 4 3 1 4 2 2 0 2 0 3 2 4 3 3 0 0 0 3 1 0 6 4 1 1 3 0 3 3 4 5 4 0 1
## [1074] 3 2 4 3 2 0 0 3 0 4 1 1 2 4 0 3 5 4 3 0 3 4 0 2 1 2 1 0 0 0 1 0 2 3 2 4 1
## [1111] 3 3 0 0 1 4 3 0 0 3 4 2 0 0 2 0 0 0 2 0 1 0 1 4 2 0 4 0 0 3 3 0 3 1 0 4 0
## [1148] 4 0 1 1 0 0 4 3 0 0 0 2 0 1 4 1 4 2 0 0 1 2 0 4 2 2 2 4 4 0 3 1 4 0 1 0 0
## [1185] 5 2 4 0 4 0 1 3 1 2 0 0 0 3 1 3 0 4 4 1 0 1 3 3 1 0 3 5 0 3 4 3 4 0 4 0 0
## [1222] 5 2 4 3 3 0 0 3 0 3 3 0 2 0 0 3 0 0 1 0 2 3 2 3 0 4 3 2 2 0 0 2 1 1 0 0 0
## [1259] 0 0 0 1 3 2 3 0 0 3 1 1 4 3 0 0 4 3 3 3 3 0 3 0 3 3 0 1 0 3 0 2 2 0 0 1 4
## [1296] 0 0 1 2 0 1 1 1 0 1 2 2 1 2 0 4 1 0 4 2 0 4 1 0 0 4 0 3 0 4 2 1 2 3 0 2 3
## [1333] 0 0 0 0 2 1 4 1 1 3 0 3 0 0 4 0 1 4 3 2 3 1 0 0 3 3 1 1 0 0 0 2 3 3 2 3 0
## [1370] 2 0 2 1 0 0 2 0 0 0 0 4 3 0 0 1 1 0 1 0 0 1 0 1 2 3 1 0 0 0 4 1 1 0 3 3 3
## [1407] 1 2 2 1 2 3 4 2 3 3 0 4 2 3 1 4 2 3 0 0 0 0 5 5 2 3 0 0 1 0 0 0 2 0 3 0 0
## [1444] 0 2 0 0 3 5 0 3 0 2 0 3 3 4 0 0 3 0 2 0 4 3 0 0 0 0 1 0 2 0 1 3 0 3 1 4 2
## [1481] 2 2 3 4 1 0 3 0 3 1 1 1 4 1 0 2 0 1 0 2 0 5 3 4 0 0 1 0 5 3 1 3 4 0 0 0 0
## [1518] 5 1 2 4 2 2 4 3 4 5 0 3 0 5 1 0 1 2 3 4 4 2 4 1 2 5 5 3 0 0 3 1 1 1 3 1 1
## [1555] 0 0 3 2 3 0 2 3 3 2 1 2 0 0 4 4 3 0 1 3 3 2 3 1 1 3 0 0 1 4 3 1 2 0 1 5 0
## [1592] 2 0 0 3 0 2 2 0 0 1 2 1 4 2 0 0 3 2 2 1 3 0 0 1 0 4 3 1 1 3 4 4 2 2 1 0 4
## [1629] 3 4 0 1 0 3 3 3 4 1 4 0 1 3 0 3 2 2 4 0 1 5 0 3 2 4 4 0 4 4 5 0 3 2 1 2 4
```

```
## [1666] 0 0 1 2 2 1 0 4 4 0 0 3 3 5 0 0 4 0 0 0 3 3 0 4 0 3 3 0 3 3 0 0 1 0 0 3 1
## [1703] 1 0 3 0 0 4 2 0 3 2 0 0 3 4 2 1 0 1 1 2 4 2 0 1 4 3 0 3 0 2 4 1 2 3 0 3 1
## [1740] 1 4 5 5 3 2 4 4 2 5 0 3 1 0 3 1 5 4 2 0 2 2 0 2 2 0 1 2 4 2 2 0 2 0 0 0 0
## [1777] 1 1 3 2 1 2 0 0 2 0 3 4 1 4 4 1 0 0 3 0 3 1 4 1 3 2 0 0 3 1 3 1 0 4 0 4 3
## [1814] 3 2 1 3 3 5 3 0 3 0 3 0 2 2 0 0 0 3 0 1 3 0 4 0 5 0 4 2 1 4 1 2 4 2 0 4 5
## [1851] 0 0 0 0 0 3 3 1 5 0 1 0 0 0 1 4 0 0 3 3 0 4 0 1 2 1 4 4 0 0 0 0 2 1 2 1 0
## [1888] 2 3 3 3 4 0 4 2 4 0 3 0 3 0 0 1 0 1 2 0 2 3 0 1 3 3 0 1 3 0 3 0 2 3 2 0 0
## [1925] 3 2 3 0 0 1 4 3 1 5 4 0 2 1 2 4 0 4 3 5 1 3 2 2 3 0 2 0 2 2 1 1 1 3 0 3 0
## [1962] 1 2 0 3 3 0 2 5 0 0 1 4 4 0 1 0 5 0 3 0 1 4 3 1 0 2 0 2 0 0 2 0 0 3 3 0 3
## [1999] 4 0 0 3 1 3 4 2 3 0 0 1 5 4 0 2 4 1 4 3 0 0 0 0 0 1 4 0 4 0 3 4 0 2 0 2 4
## [2036] 0 1 1 1 2 0 2 0 3 2 0 0 3 4 4 1 2 0 3 0 3 4 4 4 0 2 1 2 4 2 4 3 1 4 4 1 0
## [2073] 3 2 4 0 4 3 0 2 5 0 0 0 1 3 6 1 3 0 1 0 3 3 1 0 0 0 3 3 4 1 1 1 1 3 5 0 5
## [2110] 0 1 0 0 1 0 0 0 1 0 0 1 1 4 0 1 5 2 2 0 3 0 0 0 0 0 0 1 4 4 4 2 1 0 0 1 4
## [2147] 0 0 3 2 2 2 0 2 1 3 1 5 0 0 3 1 3 3 3 3 4 3 1 0 0 0 1 0 1 0 0 3 4 1 2 3 0
## [2184] 3 3 0 3 0 0 2 0 5 3 2 0 1 1 3 3 1 1 0 0 3 0 1 1 1 1 0 2 2 0 1 0 4 0 0 0 4
## [2221] 3 3 1 2 0 0 4 1 0 0 4 1 1 0 0 2 3 3 3 0 0 4 5 1 0 4 3 0 0 0 3 0 1 3 3 1 2
## [2258] 0 0 0 0 1 3 4 1 3 0 0 0 2 5 3 0 0 0 0 0 0 4 0 1 2 1 0 4 3 0 4 3 4 1 3 4 1
## [2295] 2 2 1 1 3 4 4 5 3 4 3 4 2 1 1 2 3 4 0 2 0 1 2 0 2 2 1 0 4 0 0 0 2 1 0 0 4
## [2332] 3 1 0 0 4 0 1 0 4 0 3 1 1 2 0 1 3 3 1 2 2 3 2 4 0 4 4 2 1 0 3 0 0 2 0 0 0
## [2369] 1 2 0 0 2 0 2 2 3 4 3 1 1 1 2 3 0 0 0 0 1 2 1 0 0 4 3 4 0 0 2 0 1 0 0 4 3
## [2406] 0 0 3 1 3 0 0 4 0 4 0 0 0 0 1 5 2 0 1 1 1 2 1 3 5 3 3 0 1 0 0 2 3 0 3 2 4
## [2443] 3 0 0 3 2 0 0 2 3 2 1 0 2 3 0 3 0 4 0 1 2 0 2 0 1 3 2 3 4 3 2 3 0 3 4 0 3
## [2480] 0 1 4 4 1 0 4 0 0 4 0 2 4 0 1 0 2 0 0 2 0 2 0 4 1 4 2 0 3 3 1 0 0 3 3 0 5
## [2517] 2 3 2 4 3 2 1 2 3 0 1 3 0 3 2 0 0 1 0 1 0 0 3 3 2 3 0 0 0 1 0 4 2 0 2 1 1
## [2554] 4 2 2 0 1 1 0 1 3 2 1 0 0 0 1 0 0 0 0 4 0 5 0 0 1 3 2 5 1 1 2 0 0 1 5 1 2
## [2591] 1 0 0 2 1 2 4 0 0 1 0 2 3 3 4 0 3 2 1 4 3 2 3 0 3 1 4 0 2 0 4 0 0 0 2 2 0
## [2628] 0 2 1 1 5 3 0 4 0 2 3 1 0 3 3 4 1 0 2 3 2 1 3 2 0 1 3 2 1 3 4 3 0 2 2 0 3
## [2665] 4 0 0 1 0 0 0 4 2 1 0 0 0 1 0 5 1 2 0 1 1 1 0 0 0 2 2 0 1 1 2 3 4 1 0 0 2
## [2702] 1 3 3 0 3 4 3 0 0 3 3 4 0 0 3 4 0 3 0 2 0 1 1 0 3 3 0 0 5 2 3 3 1 2 5 0 5
## [2739] 0 3 2 1 4 4 3 1 4 0 0 3 0 1 3 4 3 3 2 4 2 3 2 0 0 0 4 0 0 0 4 5 0 2 0 0 2
## [2776] 1 0 3 1 2 1 4 2 3 0 0 3 3 0 0 4 0 0 1 4 5 3 0 2 5 0 3 0 3 3 1 2 0 0 4 3 0
## [2813] 3 2 2 2 3 0 2 0 2 0 5 2 1 3 0 2 3 1 1 1 0 2 0 3 0 4 4 1 3 0 1 0 1 5 2 4 1
## [2850] 4 3 4 4 0 0 2 3 0 2 3 3 3 2 0 2 0 3 0 0 4 4 3 0 1 0 2 0 0 0 0 0 4 0 1 5 0
## [2887] 4 2 1 1 0 0 2 2 2 0 1 2 1 4 2 1 4 4 4 4 2 1 3 0 4 0 3 3 1 4 3 6 2 3 3 1 0
## [2924] 0 0 1 0 3 0 4 0 0 3 2 4 4 4 4 0 0 3 1 0 0 0 2 2 0 1 2 3 0 2 2 2 3 5 2 2 0
## [2961] 2 0 0 0 3 1 0 5 3 2 1 4 1 0 4 4 2 0 4 0 2 1 0 0 2 0 0 4 4 0 0 3 2 2 0 2 2
## [2998] 0 3 1 0 3 0 0 1 4 0 2 1 3 2 4 0 1 2 0 1 3 3 0 5 0 5 2 0 3 3 2 1 3 3 4 2 0
## [3035] 2 0 4 4 3 3 1 0 4 3 0 2 0 0 3 0 5 1 4 1 2 0 0 1 3 0 3 0 0 3 4 0 4 0 0 0 3
## [3072] 0 0 4 1 0 0 2 4 0 0 0 0 0 3 3 4 2 2 0 4 1 4 4 3 4 0 3 0 4 3 4 1 3 0 0 2 1
## [3109] 1 0 4 0 2 1 4 3 3 4 2 2 0 0 1 6 1 1 0 3 4 1 0 0 3 0 0 0 0 2 4 1 2 0 3 0 0
## [3146] 2 3 0 3 4 4 4 0 0 4 2 0 3 0 2 0 1 2 4 0 0 4 1 4 0 4 0 4 4 3 3 2 2 3 2 0 0
## [3183] 3 0 3 0 0 3 4 3 2 2 4 0 0 0 3 3 3 2 0 3 0 1 1 4 3 0 1 2 1 0 2 3 3 0 3 1 0
## [3220] 1 1 4 0 0 0 2 0 0 4 5 0 0 1 4 0 2 0 3 0 3 0 2 2 2 0 1 0 1 4 1 0 0 0 4 1 2
## [3257] 3 0 4 1 0 1 4 4 0 0 2 0 4 3 2 2 4 4 1 0 0 5 3 5 1 4 0 1 4 0 0 0 2 1 3 3 0
## [3294] 4 0 4 3 3 3 0 2 0 3 4 3 1 4 3 2 4 1 0 0 0 0 0 2 4 3 4 2 2 5 2 2 0 1 2 0 3
## [3331] 0 3 0 1 4 0 0 0 2 4 1 0 0 0 2 0 3 2 1 0 4 4 1 0 3 3 3 0 2 5 2 0 1 0 0 4 1
## [3368] 1 0 0 0 2 0 0 3 4 0 0 1 3 2 0 0 3 0 0 3 5 3 1 0 3 2 4 0 0 5 2 2 2 0 0 0 2
## [3405] 4 0 2 4 1 4 4 4 0 0 0 0 4 3 0 2 0 1 4 3 1 1 3 4 3 1 4 1 4 3 0 0 0 3 0 4 0
## [3442] 0 5 3 0 0 3 4 6 0 3 2 2 0 2 0 1 0 0 3 2 4 0 3 2 0 2 4 3 0 0 0 1 1 4 3 2 2
## [3479] 1 0 0 2 3 4 0 5 2 2 0 1 3 2 3 4 0 0 5 1 2 3 1 3 3 2 0 0 0 3 3 0 3 4 0 1 0
## [3516] 0 4 2 2 1 0 0 1 3 2 1 0 4 2 0 2 3 0 0 3 0 3 1 1 2 0 1 4 0 0 2 1 0 2 0 0 3
## [3553] 2 0 2 4 2 3 1 0 2 1 3 0 3 1 3 0 0 3 1 0 0 0 2 3 4 2 0 3 1 4 1 0 3 0 0 0 3
## [3590] 0 1 3 2 3 2 1 0 3 3 2 0 0 3 2 0 0 2 0 0 0 3 0 2 3 0 2 2 0 1 1 4 1 0 2 3 2
## [3627] 2 0 0 4 0 0 3 3 0 3 2 0 4 1 0 4 2 0 0 0 1 0 3 2 0 1 2 1 4 0 0 2 3 4 3 0 4
```

```
## [3664] 4 2 5 0 0 5 0 4 3 4 3 3 0 4 4 0 0 3 0 0 0 3 1 0 1 0 1 2 1 0 0 0 4 0 0 1 4
## [3701] 5 0 4 4 4 5 0 2 0 1 1 0 0 0 2 0 0 0 0 0 0 4 0 0 3 2 0 0 5 0 1 1 0 1 3 2 0
## [3738] 0 0 3 2 2 3 1 2 3 3 4 0 3 0 2 1 1 0 4 3 1 1 2 0 1 0 2 0 3 0 4 0 3 2 1 0 0
## [3775] 0 2 0 1 3 1 0 0 3 0 2 3 1 0 5 0 4 1 2 1 0 3 0 1 1 2 4 3 4 2 0 2 5 4 4 3 0
## [3812] 4 2 0 0 0 1 0 3 0 0 1 1 4 1 3 3 0 1 5 0 1 0 2 4 1 0 1 3 0 0 2 0 0 2 1 0 0
## [3849] 1 2 1 3 0 1 1 3 5 4 2 4 3 0 0 3 1 0 1 3 1 0 3 0 0 3 0 0 1 0 0 0 2 3 4 3 0
## [3886] 0 3 4 4 0 0 3 1 2 3 2 0 3 0 3 1 0 1 0 3 2 4 0 3 2 1 0 3 4 1 2 3 1 2 2 3 1
## [3923] 0 0 1 0 3 0 0 2 2 3 0 4 0 4 1 0 0 0 0 1 2 0 5 2 3 3 0 2 4 2 2 2 1 1 0 2 2
## [3960] 1 2 4 0 1 0 1 1 0 1 1 0 0 2 0 3 3 2 4 0 1 4 3 0 1 0 2 3 2 2 0 0 1 4 3 1 4
## [3997] 1 3 5 3 2 0 1 0 1 4 2 0 1 2 0 4 0 3 0 3 1 4 0 0 2 0 0 3 3 3 0 0 3 0 3 2 0
## [4034] 0 0 3 0 3 0 0 1 0 3 4 1 0 0 0 4 0 2 0 0 1 0 0 4 2 0 4 2 0 0 5 1 4 2 3 0 1
## [4071] 5 1 3 3 0 1 3 1 1 1 4 3 3 0 2 3 1 5 0 1 0 0 0 1 4 2 2 4 4 3 2 0 0 4 3 1 2
## [4108] 1 0 2 0 1 3 4 0 2 0 0 3 3 3 2 3 0 1 0 0 1 3 1 4 3 0 2 1 5 2 2 5 5 3 3 0 0
## [4145] 3 1 3 0 2 1 4 1 0 0 0 0 3 0 4 0 0 3 2 1 4 3 0 1 2 2 4 2 1 0 0 0 3 0 0 3 0
## [4182] 3 3 2 3 2 1 3 5 4 0 0 0 1 3 0 0 4 0 0 3 1 1 0 0 0 0 4 0 0 1 0 3 2 1 0 3 3
## [4219] 1 0 3 1 1 3 0 0 1 0 2 0 0 3 3 3 4 3 2 1 1 1 0 0 0 4 0 2 0 0 0 0 0 0 0 3 2 2
## [4256] 3 1 4 3 0 4 2 1 3 1 0 0 3 4 3 0 0 0 4 4 1 2 0 3 3 0 4 2 2 2 2 2 0 0 2 4 2
## [4293] 3 2 3 0 2 0 0 0 0 0 2 4 2 0 0 0 4 0 3 0 2 0 0 1 3 1 3 0 0 1 6 0 0 1 1 4 3
## [4330] 1 0 1 3 2 1 3 1 1 2 0 2 2 1 4 2 0 3 2 0 2 3 3 0 3 1 4 2 0 3 0 3 3 3 0 2 1
## [4367] 4 2 0 4 0 3 0 3 0 0 3 0 0 3 1 0 4 0 1 1 0 2 1 2 3 0 0 0 2 3 2 2 0 3 0 0 0
## [4404] 2 3 3 1 1 1 2 0 1 3 3 3 4 0 1 1 0 0 3 0 1 0 0 2 1 4 4 4 0 4 1 3 0 0 0 0 2
## [4441] 4 6 2 1 0 2 2 0 2 0 2 4 0 1 2 0 0 0 1 1 3 0 0 0 0 3 3 1 1 2 2 0 0 4 3 3 3
## [4478] 3 0 2 0 0 0 1 0 3 0 0 0 0 0 1 1 0 4 0 0 2 0 3 3 0 3 4 2 3 0 2 1 3 4 5 0 1
## [4515] 0 3 4 1 0 0 0 0 2 3 1 4 6 0 3 0 0 0 0 4 0 0 0 0 3 3 0 3 1 2 3 1 3 1 0 2 1
## [4552] 0 4 0 2 0 3 2 0 1 3 2 1 2 2 3 2 3 1 1 2 3 2 0 4 0 0 0 1 1 0 0 0 0 0 0 0 0
## [4589] 0 3 3 3 4 1 4 3 2 1 1 0 2 5 3 1 2 0 0 0 0 4 0 0 1 3 0 0 0 1 1 3 3 0 0 1 0
## [4626] 3 0 3 1 1 2 3 1 0 4 0 4 2 0 0 2 4 1 4 1 2 0 0 0 0 1 4 0 0 1 0 1 0 2 4 2 0
## [4663] 0 0 4 0 3 3 0 3 1 0 3 0 4 0 3 0 1 0 5 1 4 4 2 3 2 1 4 0 0 0 2 0 5 4 0 5 2
## [4700] 0 0 3 4 0 2 0 0 4 0 3 0 0 1 4 2 3 0 4 1 1 2 0 6 0 4 4 1 1 1 0 3 3 3 3 0 1
## [4737] 0 1 1 4 2 0 3 3 0 0 3 4 0 0 1 0 1 2 2 5 0 4 2 4 0 4 2 1 1 2 1 3 0 0 4 0 0
## [4774] 0 1 3 0 2 0 5 2 3 1 0 0 2 2 0 2 0 3 3 1 0 0 4 2 2 1 2 3 4 0 0 0 2 1 1 3 2
## [4811] 3 1 0 3 0 2 0 4 3 3 1 5 0 2 2 2 2 3 0 2 4 1 0 1 3 3 2 2 2 2 0 1 0 5 3 0 0
## [4848] 3 2 4 1 0 3 0 3 0 0 1 1 3 3 0 4 0 4 0 0 0 0 0 2 0 3 1 4 4 0 0 2 1 0 1 1 1
## [4885] 0 4 2 0 4 4 0 3 0 3 0 2 0 4 0 4 3 1 3 1 0 2 4 2 3 3 0 3 0 3 1 2 5 1 4 0 0
## [4922] 0 0 3 0 0 2 1 0 0 0 3 3 0 0 3 0 2 1 0 3 3 3 0 4 2 4 0 4 4 0 0 0 2 4 1 0 3
## [4959] 0 2 1 5 2 1 0 0 0 0 4 0 3 0 0 0 2 0 1 3 0 4 4 0 0 2 4 0 2 3 3 2 1 4 1 2 3
## [4996] 1 0 1 3 4 3 4 4 2 3 4 0 1 0 1 4 3 1 1 5 0 1 3 3 1 1 1 0 0 0 0 0 4 2 4 3 2
## [5033] 0 0 3 4 1 0 0 2 0 3 0 0 1 2 1 1 4 0 4 0 0 0 2 0 1 0 0 3 1 1 1 3 5 3 2 0 4
## [5070] 3 0 0 3 0 0 4 4 3 0 1 2 0 4 3 2 5 1 4 4 1 2 0 1 0 0 1 3 0 0 2 0 0 3 1 4 2
## [5107] 3 1 0 0 0 0 4 3 3 3 1 0 2 4 0 3 2 2 0 0 0 0 4 0 4 0 0 0 0 3 0 1 1 4 1 2 1
## [5144] 1 1 2 3 2 3 2 0 1 2 5 0 3 3 0 0 0 0 1 0 1 1 3 0 3 1 3 2 0 0 0 3 0 2 0 1 1
## [5181] 4 3 0 0 0 3 0 0 2 0 0 0 4 2 3 3 2 0 2 2 3 0 3 1 0 2 1 3 3 3 2 0 3 3 1 5 0
## [5218] 2 3 2 0 0 0 1 4 3
```

```
hist(tibb$checks)
```




```
?hist
```

```
## starting httpd help server ... done
```

```
library(psych)
```

```
##
## Attaching package: 'psych'
```

```
## The following objects are masked from 'package:ggplot2':
##
##     %+%, alpha
```

```
skew(tibb$checks) #data seems negatively skewed, although its ordinal
```

```
## [1] 0.4299042
```

```
tibb.5mean<-mean(tibb$checks[1:5])
tibb.5mean
```

```
## [1] 2.2
```

```r
tibb.mean<-mean(tibb$checks)
tibb.mean
```

```
## [1] 1.636433
```

```r
tibb.med<-median(tibb$checks)
tibb.med #since ordinal, expected to be a whole no.
```

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

```r
tibb.mode<-modeOf(tibb$checks)
tibb.mode #most frquently occuring value, validated by histogram
```

```
## [1] 0
```

```r
tibb.mode2<-modeOf(tibb$age)
tibb.mode2
```

```
## [1] 18
```

```r
dataset <- c(-15,2,3,5,6,8,9,12)
meanyes<-mean(x = dataset, trim = .13)
mean<-mean(dataset)
meanno<-mean(x=dataset, trim=0.12)
meanyes
```

```
## [1] 5.5
```

```r
mean
```

```
## [1] 3.75
```

```r
meanno
```

```
## [1] 3.75
```

```r
#Try calculating 12% trimmed mean for above dataset - does not do anything, since 12 percent of 8 value
```

```r
tibb.range<-range(tibb$checks)
tibb.range
```

```
## [1] 0 6
```

```r
rangedata<-range(dataset)
rangedata
```

```
## [1] -15  12
```

```
#range() not giving difference, it is the same range used in assignment 1
```

```
tibb2<-read_csv("C:/Users/ushad/Downloads/SmokeBan.csv")
```

```
## Rows: 10000 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (6): smoker, ban, education, afam, hispanic, gender
## dbl (2): rownames, age
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
tibb2
```

```
## # A tibble: 10,000 x 8
##    rownames smoker ban     age education    afam  hispanic gender
##       <dbl> <chr>  <chr> <dbl> <chr>        <chr> <chr>    <chr>
##  1        1 yes    yes      41 hs           no    no       female
##  2        2 yes    yes      44 some college no    no       female
##  3        3 no     no       19 some college no    no       female
##  4        4 yes    no       29 hs           no    no       female
##  5        5 no     yes      28 some college no    no       female
##  6        6 no     no       40 some college no    no       male
##  7        7 yes    yes      47 some college no    no       female
##  8        8 yes    no       36 some college no    no       male
##  9        9 no     yes      49 some college no    no       female
## 10       10 no     no       44 some college no    no       male
## # i 9,990 more rows
```

```
newquant<-quantile(x = tibb2$age, prob = 0.25)
newquant2<-quantile(x = tibb2$age, prob = 0.75)
newquant
```

```
## 25%
##  29
```

```
newquant2
```

```
## 75%
##  47
```

```
intqra<-IQR(tibb2$age)
nq3<-newquant2-newquant
nq3 #why does this show 75% with the value (like newquant and newquant2)?
```

```
## 75%
##  18
```

```r
mode(nq3)
```

```
## [1] "numeric"
```

```r
intqra
```

```
## [1] 18
```

```r
vari<-var(tibb2$age)
stddev<-sd(tibb2$age)
vari
```

```
## [1] 146.7437
```

```r
stddev
```

```
## [1] 12.11378
```

```r
vari^(1/2)
```

```
## [1] 12.11378
```

```r
summary(tibb2$age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   29.00   37.00   38.69   47.00   88.00
```

```r
var(tibb$checks)
```

```
## [1] 2.369229
```

```r
sd(tibb$checks)
```

```
## [1] 1.53923
```

```r
summary(tibb$checks)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   1.000   1.636   3.000   6.000
```

```r
?who  #The who function prints out some basic information about all variables in the workspace.
```

```r
describeBy(x=tibb2)
```

```
## Warning in describeBy(x = tibb2): no grouping variable requested
```

```
##            vars     n    mean      sd median trimmed     mad min   max range
## rownames      1 10000 5000.50 2886.90 5000.5 5000.50 3706.50   1 10000  9999
## smoker*       2 10000    1.24    0.43    1.0    1.18    0.00   1     2     1
## ban*          3 10000    1.61    0.49    2.0    1.64    0.00   1     2     1
## age           4 10000   38.69   12.11   37.0   37.98   13.34  18    88    70
## education*    5 10000    2.94    1.53    2.0    2.93    1.48   1     5     4
## afam*         6 10000    1.08    0.27    1.0    1.00    0.00   1     2     1
## hispanic*     7 10000    1.11    0.32    1.0    1.02    0.00   1     2     1
## gender*       8 10000    1.44    0.50    1.0    1.42    0.00   1     2     1
##              skew kurtosis    se
## rownames    0.00    -1.20 28.87
## smoker*     1.20    -0.55  0.00
## ban*       -0.45    -1.80  0.00
## age         0.51    -0.22  0.12
## education*  0.23    -1.50  0.02
## afam*       3.18     8.08  0.00
## hispanic*   2.44     3.94  0.00
## gender*     0.26    -1.93  0.00
```

```
describeBy(x=tibb2, group = tibb2$education)
```

```
##
##  Descriptive statistics by group
## group: college
##            vars    n    mean      sd median trimmed     mad min  max range  skew
## rownames      1 1972 4847.02 2941.92 4753.5 4811.74 3965.96  27 9988  9961  0.09
## smoker        2 1972    1.14    0.34    1.0    1.04    0.00   1    2     1  2.12
## ban           3 1972    1.69    0.46    2.0    1.74    0.00   1    2     1 -0.84
## age           4 1972   37.87   10.63   36.0   37.04   11.86  20   79    59  0.62
## education     5 1972    1.00    0.00    1.0    1.00    0.00   1    1     0   NaN
## afam          6 1972    1.05    0.22    1.0    1.00    0.00   1    2     1  3.98
## hispanic      7 1972    1.05    0.22    1.0    1.00    0.00   1    2     1  4.07
## gender        8 1972    1.47    0.50    1.0    1.46    0.00   1    2     1  0.12
##            kurtosis    se
## rownames      -1.25 66.25
## smoker         2.51  0.01
## ban           -1.30  0.01
## age           -0.18  0.24
## education       NaN  0.00
## afam          13.82  0.01
## hispanic      14.56  0.00
## gender        -1.99  0.01
## ------------------------------------------------------------
## group: hs
##            vars    n    mean      sd median trimmed     mad min  max range  skew
## rownames      1 3266 5505.33 2755.99   5840 5588.24 3222.43   1 9999  9998 -0.21
## smoker        2 3266    1.31    0.46      1    1.27    0.00   1    2     1  0.80
## ban           3 3266    1.56    0.50      2    1.57    0.00   1    2     1 -0.22
## age           4 3266   39.15   12.77     38   38.50   14.83  18   78    60  0.43
## education     5 3266    2.00    0.00      2    2.00    0.00   2    2     0   NaN
## afam          6 3266    1.09    0.28      1    1.00    0.00   1    2     1  2.94
## hispanic      7 3266    1.10    0.30      1    1.00    0.00   1    2     1  2.68
## gender        8 3266    1.38    0.48      1    1.35    0.00   1    2     1  0.50
##            kurtosis    se
```

```
## rownames     -1.04 48.22
## smoker       -1.36  0.01
## ban          -1.95  0.01
## age          -0.49  0.22
## education      NaN  0.00
## afam          6.63  0.00
## hispanic      5.19  0.01
## gender       -1.75  0.01
## ------------------------------------------------------------
## group: hs drop out
##           vars   n    mean      sd median trimmed     mad min  max range  skew
## rownames     1 912 4667.51 2903.18   4721 4622.66 3763.58  26 9993  9967  0.08
## smoker       2 912    1.38    0.48      1    1.35    0.00   1    2     1  0.51
## ban          3 912    1.46    0.50      1    1.45    0.00   1    2     1  0.15
## age          4 912   40.38   14.12     39   39.83   16.31  18   81    63  0.33
## education    5 912    3.00    0.00      3    3.00    0.00   3    3     0   NaN
## afam         6 912    1.09    0.29      1    1.00    0.00   1    2     1  2.84
## hispanic     7 912    1.40    0.49      1    1.38    0.00   1    2     1  0.39
## gender       8 912    1.54    0.50      2    1.55    0.00   1    2     1 -0.15
##           kurtosis    se
## rownames     -1.23 96.13
## smoker       -1.74  0.02
## ban          -1.98  0.02
## age          -0.69  0.47
## education      NaN  0.00
## afam          6.07  0.01
## hispanic     -1.85  0.02
## gender       -1.98  0.02
## ------------------------------------------------------------
## group: master
##           vars    n    mean      sd median trimmed     mad min  max range  skew
## rownames     1 1048 4630.12 2787.78   4199 4566.74 3444.82  24 9997  9973  0.18
## smoker       2 1048    1.09    0.29      1    1.00    0.00   1    2     1  2.81
## ban          3 1048    1.72    0.45      2    1.77    0.00   1    2     1 -0.97
## age          4 1048   42.17   10.20     41   41.53   10.38  22   88    66  0.65
## education    5 1048    4.00    0.00      4    4.00    0.00   4    4     0   NaN
## afam         6 1048    1.04    0.20      1    1.00    0.00   1    2     1  4.68
## hispanic     7 1048    1.03    0.17      1    1.00    0.00   1    2     1  5.55
## gender       8 1048    1.55    0.50      2    1.56    0.00   1    2     1 -0.19
##           kurtosis    se
## rownames     -1.12 86.11
## smoker        5.89  0.01
## ban          -1.06  0.01
## age           0.47  0.32
## education      NaN  0.00
## afam         19.95  0.01
## hispanic     28.78  0.01
## gender       -1.97  0.02
## ------------------------------------------------------------
## group: some college
##           vars    n    mean      sd median trimmed     mad min   max range
## rownames     1 2802 4767.00 2950.47 4612.5 4725.23 3889.60   2 10000  9998
## smoker       2 2802    1.25    0.43    1.0    1.18    0.00   1     2     1
## ban          3 2802    1.62    0.49    2.0    1.65    0.00   1     2     1
```

```
## age          4 2802   36.90   11.89   35.0   36.00   11.86  18    88    70
## education    5 2802    5.00    0.00    5.0    5.00    0.00   5     5     0
## afam         6 2802    1.09    0.29    1.0    1.00    0.00   1     2     1
## hispanic     7 2802    1.11    0.31    1.0    1.01    0.00   1     2     1
## gender       8 2802    1.41    0.49    1.0    1.38    0.00   1     2     1
##            skew kurtosis    se
## rownames   0.10    -1.24 55.74
## smoker     1.18    -0.61  0.01
## ban       -0.50    -1.75  0.01
## age        0.65     0.08  0.22
## education   NaN      NaN  0.00
## afam       2.84     6.04  0.01
## hispanic   2.48     4.16  0.01
## gender     0.38    -1.86  0.01
```

```
#we get many dataframes, each describes the summary for a particular type of education, their n's would
```

```
summary(tibb2)
```

```
##     rownames        smoker              ban                 age
##  Min.   :    1   Length:10000       Length:10000       Min.   :18.00
##  1st Qu.: 2501   Class :character   Class :character   1st Qu.:29.00
##  Median : 5000   Mode  :character   Mode  :character   Median :37.00
##  Mean   : 5000                                         Mean   :38.69
##  3rd Qu.: 7500                                         3rd Qu.:47.00
##  Max.   :10000                                         Max.   :88.00
##   education             afam             hispanic            gender
##  Length:10000       Length:10000       Length:10000       Length:10000
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```
by(data=tibb2, INDICES=tibb2$gender, FUN=summary) #gives summary for each gender separately, we can for
```

```
## tibb2$gender: female
##     rownames       smoker              ban                 age
##  Min.   :   1   Length:5637        Length:5637        Min.   :18.00
##  1st Qu.:2642   Class :character   Class :character   1st Qu.:30.00
##  Median :5110   Mode  :character   Mode  :character   Median :38.00
##  Mean   :5081                                         Mean   :38.99
##  3rd Qu.:7533                                         3rd Qu.:47.00
##  Max.   :9999                                         Max.   :88.00
##   education             afam             hispanic            gender
##  Length:5637        Length:5637        Length:5637        Length:5637
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
## ------------------------------------------------------------
```

```
## tibb2$gender: male
##     rownames        smoker              ban                 age
##  Min.   :    6   Length:4363        Length:4363        Min.   :18.00
##  1st Qu.: 2318   Class :character   Class :character   1st Qu.:29.00
##  Median : 4889   Mode  :character   Mode  :character   Median :37.00
##  Mean   : 4897                                         Mean   :38.31
##  3rd Qu.: 7462                                         3rd Qu.:46.00
##  Max.   :10000                                         Max.   :88.00
##   education             afam             hispanic            gender
##  Length:4363        Length:4363        Length:4363        Length:4363
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```
by(data=tibb2, INDICES=tibb2$gender, FUN=describe)
```

```
## tibb2$gender: female
##            vars    n    mean      sd median trimmed     mad min  max range
## rownames      1 5637 5080.95 2855.99   5110 5095.57 3621.99   1 9999  9998
## smoker*       2 5637    1.23    0.42      1    1.16    0.00   1    2     1
## ban*          3 5637    1.66    0.47      2    1.70    0.00   1    2     1
## age           4 5637   38.99   12.16     38   38.29   13.34  18   88    70
## education*    5 5637    2.94    1.54      2    2.93    1.48   1    5     4
## afam*         6 5637    1.09    0.28      1    1.00    0.00   1    2     1
## hispanic*     7 5637    1.10    0.30      1    1.00    0.00   1    2     1
## gender*       8 5637    1.00    0.00      1    1.00    0.00   1    1     0
##             skew kurtosis    se
## rownames   -0.03    -1.17 38.04
## smoker*     1.28    -0.36  0.01
## ban*       -0.67    -1.55  0.01
## age         0.50    -0.26  0.16
## education*  0.28    -1.51  0.02
## afam*       2.96     6.74  0.00
## hispanic*   2.61     4.80  0.00
## gender*      NaN      NaN  0.00
## ------------------------------------------------------------
## tibb2$gender: male
##            vars    n    mean      sd median trimmed     mad min   max range
## rownames      1 4363 4896.55 2923.40   4889 4876.62 3813.25   6 10000  9994
## smoker*       2 4363    1.26    0.44      1    1.20    0.00   1     2     1
## ban*          3 4363    1.55    0.50      2    1.56    0.00   1     2     1
## age           4 4363   38.31   12.05     37   37.60   13.34  18    88    70
## education*    5 4363    2.95    1.52      3    2.93    1.48   1     5     4
## afam*         6 4363    1.07    0.25      1    1.00    0.00   1     2     1
## hispanic*     7 4363    1.13    0.33      1    1.03    0.00   1     2     1
## gender*       8 4363    1.00    0.00      1    1.00    0.00   1     1     0
##             skew kurtosis    se
## rownames    0.04    -1.24 44.26
## smoker*     1.11    -0.77  0.01
## ban*       -0.18    -1.97  0.01
## age         0.53    -0.17  0.18
## education*  0.17    -1.48  0.02
```

```
## afam*       3.52    10.37  0.00
## hispanic*   2.25     3.05  0.01
## gender*      NaN      NaN  0.00
```

```
#gives 2 dataframes as output, each with summary in describe format for each sex
```

```
aggregate( x = age ~ gender + afam,
           data = tibb2,
           FUN = median)
```

```
##   gender afam age
## 1 female   no  38
## 2   male   no  37
## 3 female  yes  37
## 4   male  yes  37
```

```
#x is used instead of formula now
```

```
aggregate( x = age ~ afam + gender,
           data = tibb2,
           FUN = median)
```

```
##   afam gender age
## 1   no female  38
## 2  yes female  37
## 3   no   male  37
## 4  yes   male  37
```

```
#Try interchanging the positions of drug and therapy above - values are same, just order of arrangement
```

```
aggregate( x = age ~ colour + released,
           data = tibb,
           FUN = median)
```

```
##   colour released age
## 1  Black       No  22
## 2  White       No  21
## 3  Black      Yes  22
## 4  White      Yes  21
```