Example RMarkdown

Monica Alexander

10/01/2022

Hi my name is Monica blah blah

```
# this is a comment
x <- 853
y <- 90
x+y
```

```
## [1] 943
```

Read in the GSS dataset

```
#  load in tidyverse package
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
# read in the GSS
gss <- read_csv(file = "../data/gss.csv")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_character(),
##   caseid = col_double(),
##   age = col_double(),
##   age_first_child = col_double(),
##   age_youngest_child_under_6 = col_double(),
##   total_children = col_double(),
##   age_start_relationship = col_double(),
##   age_at_first_marriage = col_double(),
```

```
##   age_at_first_birth = col_double(),
##   distance_between_houses = col_double(),
##   age_youngest_child_returned_work = col_double(),
##   feelings_life = col_double(),
##   hh_size = col_double(),
##   number_total_children_intention = col_double(),
##   number_marriages = col_double(),
##   fin_supp_child_supp = col_double(),
##   fin_supp_child_exp = col_double(),
##   fin_supp_lump = col_double(),
##   fin_supp_other = col_double(),
##   is_male = col_double(),
##   main_activity = col_logical()
##   # ... with 2 more columns
## )
## i Use `spec()` for the full column specifications.
```

Look at the top of the GSS dataset

```
head(gss)
```

```
## # A tibble: 6 x 85
##   caseid   age age_first_child age_youngest_chi~ total_children age_start_relat~
##    <dbl> <dbl>           <dbl>             <dbl>          <dbl>            <dbl>
## 1      1  52.7              27                NA              1               NA
## 2      2  51.1              33                NA              5               NA
## 3      3  63.6              40                NA              5               NA
## 4      4  80                56                NA              1               NA
## 5      5  28                NA                NA              0             25.3
## 6      6  63                37                NA              2               NA
## # ... with 79 more variables: age_at_first_marriage <dbl>,
## #   age_at_first_birth <dbl>, distance_between_houses <dbl>,
## #   age_youngest_child_returned_work <dbl>, feelings_life <dbl>, sex <chr>,
## #   place_birth_canada <chr>, place_birth_father <chr>,
## #   place_birth_mother <chr>, place_birth_macro_region <chr>,
## #   place_birth_province <chr>, year_arrived_canada <chr>, province <chr>,
## #   region <chr>, pop_center <chr>, marital_status <chr>, aboriginal <chr>, ...
```

Look at the dimensions of the GSS

```
ncol(gss) # number of columns
```

```
## [1] 85
```

```
nrow(gss) # rows
```

```
## [1] 20602
```

Important functions

Select columns

Let's look at column names. The function colnames returns a vector of length 85, showing us every column name

```
# column names
colnames(gss)
```

```
##  [1] "caseid"                          "age"
##  [3] "age_first_child"                 "age_youngest_child_under_6"
##  [5] "total_children"                  "age_start_relationship"
##  [7] "age_at_first_marriage"           "age_at_first_birth"
##  [9] "distance_between_houses"         "age_youngest_child_returned_work"
## [11] "feelings_life"                   "sex"
## [13] "place_birth_canada"              "place_birth_father"
## [15] "place_birth_mother"              "place_birth_macro_region"
## [17] "place_birth_province"            "year_arrived_canada"
## [19] "province"                        "region"
## [21] "pop_center"                      "marital_status"
## [23] "aboriginal"                      "vis_minority"
## [25] "age_immigration"                 "landed_immigrant"
## [27] "citizenship_status"              "education"
## [29] "own_rent"                        "living_arrangement"
## [31] "hh_type"                         "hh_size"
## [33] "partner_birth_country"           "partner_birth_province"
## [35] "partner_vis_minority"            "partner_sex"
## [37] "partner_education"               "average_hours_worked"
## [39] "worked_last_week"                "partner_main_activity"
## [41] "self_rated_health"               "self_rated_mental_health"
## [43] "religion_has_affiliation"        "religion_importance"
## [45] "language_home"                   "language_knowledge"
## [47] "income_family"                   "income_respondent"
## [49] "occupation"                      "childcare_regular"
## [51] "childcare_type"                  "childcare_monthly_cost"
## [53] "ever_fathered_child"             "ever_given_birth"
## [55] "number_of_current_union"         "lives_with_partner"
## [57] "children_in_household"           "number_total_children_intention"
## [59] "has_grandchildren"               "grandparents_still_living"
## [61] "ever_married"                    "current_marriage_is_first"
## [63] "number_marriages"                "religion_participation"
## [65] "partner_location_residence"      "full_part_time_work"
## [67] "time_off_work_birth"             "reason_no_time_off_birth"
## [69] "returned_same_job"               "satisfied_time_children"
## [71] "provide_or_receive_fin_supp"     "fin_supp_child_supp"
## [73] "fin_supp_child_exp"              "fin_supp_lump"
## [75] "fin_supp_other"                  "fin_supp_agreement"
## [77] "future_children_intention"       "is_male"
## [79] "main_activity"                   "age_diff"
## [81] "number_total_children_known"     "age_group"
## [83] "educ_cat"                        "partner_educ_cat"
## [85] "has_bachelor_or_higher"
```

Let's select the age variable

```
select(gss,age)
```

```
## # A tibble: 20,602 x 1
##      age
##    <dbl>
##  1  52.7
##  2  51.1
##  3  63.6
##  4  80
##  5  28
##  6  63
##  7  58.8
##  8  80
##  9  63.8
## 10  25.2
## # ... with 20,592 more rows
```

The pipe

The pipe is the funny thing that is %>% – remember from the lecture, when you see a pipe, read it as "and then"

```
gss %>%
  select(age, region, occupation)
```

```
## # A tibble: 20,602 x 3
##      age region           occupation
##    <dbl> <chr>            <chr>
##  1  52.7 Quebec           Sales and service occupations
##  2  51.1 Prairie region   Trades, transport and equipment operators and related~
##  3  63.6 Ontario          <NA>
##  4  80   Prairie region   <NA>
##  5  28   Quebec           Sales and service occupations
##  6  63   Quebec           <NA>
##  7  58.8 Atlantic region  <NA>
##  8  80   Quebec           <NA>
##  9  63.8 British Columbia Business, finance, and administration occupations
## 10  25.2 Prairie region   <NA>
## # ... with 20,592 more rows
```

NOTE: the above code selects three columns, but doesn't assign/save the output. If we want to assign this output to use again, just use the back arrow syntax

```
gss_ageregionocc <- gss %>%
  select(age, region, occupation)

# look at this object
gss_ageregionocc
```

```
## # A tibble: 20,602 x 3
##      age region           occupation
##    <dbl> <chr>            <chr>
##  1  52.7 Quebec           Sales and service occupations
##  2  51.1 Prairie region   Trades, transport and equipment operators and related~
##  3  63.6 Ontario          <NA>
##  4  80   Prairie region   <NA>
##  5  28   Quebec           Sales and service occupations
##  6  63   Quebec           <NA>
##  7  58.8 Atlantic region  <NA>
##  8  80   Quebec           <NA>
##  9  63.8 British Columbia Business, finance, and administration occupations
## 10  25.2 Prairie region   <NA>
## # ... with 20,592 more rows
```