Final Project

Emely Gazarov, Ryo Iwata, Maria Ramirez

16 April, 2024

Inputting the data

1.A Load data and prepare data for analysis

```
# Reading in the data

# Predictor
## Basic demographics (age, gender, education)
## Duration of sobriety prior to treatment
raw_demographics <- read_delim("./data/demo.damon.csv",
                               delim = ",",
                               progress = FALSE,
                               show_col_types = FALSE)

## Specific SUDs (e.g., alcohol use disorder)
raw_SUD <- read_delim("./data/SUDdiagnosis.damon.csv",
                               delim = ",",
                               progress = FALSE,
                               show_col_types = FALSE)

## Social Support (MSPSS)
raw_social_support <- read_delim("./data/mspss.damon.csv",
                               delim = ",",
                               progress = FALSE,
                               show_col_types = FALSE)

## Substance Use History
raw_sub_history <- read_delim("./data/subuse.damon.csv",
                               delim = ",",
                               progress = FALSE,
                               show_col_types = FALSE)
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
problems(raw_sub_history)
```

```
## # A tibble: 9 x 5
##     row   col expected actual     file
##   <int> <int> <chr>    <chr>      <chr>
## 1  1509     7 a double "late 23"  C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C_~
```

```
## 2  1705    15 a double "30's"     C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C_~
## 3  1853     9 a double "30-40"    C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C_~
## 4  1901     8 a double "23-24"    C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C_~
## 5  1998     8 a double "15-16"    C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C_~
## 6  2001     7 a double "idk"      C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C_~
## 7  2002     8 a double "15 or 16" C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C_~
## 8  2002     9 a double "15 or 16" C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C_~
## 9  2170     7 a double " 31"      C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C_~
```

```
## AA/NA Affiliation
raw_aana_affiliation <- read_delim("./data/aana.damon.csv",
                                   delim = ",",
                                   progress = FALSE,
                                   show_col_types = FALSE)
```

```
## Warning: One or more parsing issues, call `problems()` on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
problems(raw_aana_affiliation)
```

```
## # A tibble: 15 x 5
##      row   col expected actual     file
##    <int> <int> <chr>    <chr>      <chr>
##  1   547    11 a double "100-150"  C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
##  2  1141    11 a double "12-15"    C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
##  3  1234    11 a double "100+"     C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
##  4  1509    10 a number "none"     C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
##  5  1509    11 a double "none"     C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
##  6  1705    11 a double "100 "     C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
##  7  1968    11 a double "~160"     C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
##  8  2057    10 a number "none"     C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
##  9  2057    11 a double "none"     C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
## 10  2305    11 a double "none"     C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
## 11  2311    10 a number "None"     C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
## 12  2311    11 a double "None"     C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
## 13  3817    12 a double "15-20"    C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
## 14  4465    12 a double "not sure" C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
## 15  4624    12 a double "~30"      C:/Users/Ryo Iwata/Documents/GitHub/GMS6025C~
```

```
## Stressful Life Experiences (LEC-5) lec.damon

# *The structure of this dataframe is really bonkers.
#You probably want to use the variable "toyou_total"
# which is a sum of event types that the patients endorsed as having happened to them.
# We sometimes also use "toyou_wit_total", which is a similar sum score,
# but includes events that have happened to the participant
#AND events that the participant has witnessed.

raw_stressful_life  <- read_delim("./data/lec.damon.csv",
                                   delim = ",",
                                   progress = FALSE,
                                   show_col_types = FALSE)

## Spiritual Experiences (Brief R-COPE)
```

```r
raw_spiritual <- read_delim("./data/r_cope.damon.csv",
                            delim = ",",
                            progress = FALSE,
                            show_col_types = FALSE)

## Childhood Experiences (ACE)
raw_childhood <- read_delim("./data/aces.damon.csv",
                            delim = ",",
                            progress = FALSE,
                            show_col_types = FALSE)
```

```r
# Things to predict
## Quality of Life (WHOQOL-BREF): Evaluates general, physical, psychological health, social relationshi
raw_life_quality <- read_delim("./data/QOL.damon.csv",
                            delim = ",",
                            progress = FALSE,
                            show_col_types = FALSE)

raw_commitment <- read_delim("./data/change.damon.csv",
                           delim = ",",
                           progress = FALSE,
                           show_col_types = FALSE)
## Alcohol/Drug Craving (PACS): Measures the frequency and intensity of cravings.
raw_craving <- read_delim("./data/craving.damon.csv",
                          delim = ",",
                          progress = FALSE,
                          show_col_types = FALSE)

## Treatment dropout
## demo.damon (dropout_yn)
## Length of stay in treatment
## demo.damon (days_in_tx_clean)
## Impression of Change (PGIC)
raw_change_impression <- read_delim("./data/pgic.damon.csv",
                                    delim = ",",
                                    progress = FALSE,
                                    show_col_types = FALSE)

# Other
raw_data_dictionary <- read_delim("./data/Data Dictionary.csv",
                                  delim = ",",
                                  progress = FALSE,
                                  show_col_types = FALSE)
```

```r
# str(raw_demographics)
# str(raw_SUD)
# str(raw_social_support)
# str(raw_sub_history)
# str(raw_aana_affiliation)
# str(raw_stressful_life)
# str(raw_spiritual)
# str(raw_childhood)
# str(raw_life_quality)
# str(raw_commitment)
```

```r
# str(raw_craving)
# str(raw_change_impression)
# str(raw_data_dictionary)
```

```r
# Filtering out readmission
raw_demographics |> filter(!str_detect(redcap_event_name,'readmission'))
```

```
## # A tibble: 5,302 x 8
##    record_id redcap_event_name age_today gender ed_summary         sobriety_calc
##        <dbl> <chr>                 <dbl> <chr>  <chr>                      <dbl>
##  1         1 phabaseline_arm_1        27 men    High School/GED o~            20
##  2         2 phabaseline_arm_1        41 men    Bachelor's                    14
##  3         3 phabaseline_arm_1        23 women  High School/GED o~            24
##  4         4 phabaseline_arm_1        62 women  Bachelor's                     1
##  5         5 phabaseline_arm_1        57 women  Associate's/Some ~             4
##  6         6 phabaseline_arm_1        33 women  High School/GED o~             9
##  7         7 phabaseline_arm_1        38 women  Bachelor's                     3
##  8         8 phabaseline_arm_1        26 women  High School/GED o~            13
##  9         9 phabaseline_arm_1        46 women  Associate's/Some ~           370
## 10        10 phabaseline_arm_1        23 men    Master's                       9
## # i 5,292 more rows
## # i 2 more variables: days_in_tx_clean <dbl>, dropout_yn <chr>
```

```r
# Filtering for subjects that are in each time

discharged_demo <- raw_demographics |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_demo <- raw_demographics |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_demo <- raw_demographics |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))
```

```r
joined_demographics <- baseline_demo |>
  inner_join(discharged_demo, by="record_id") |>
  inner_join(followup_demo, by="record_id")
```

Getting baseline, followup, discharge for all datasets

```r
# Filtering for subjects that are in each time

factored_SUD <- raw_SUD |> mutate(sa_dx___0 = as_factor(sa_dx___0)) |>
  mutate(sa_dx___1 = as_factor(sa_dx___1)) |>
  mutate(sa_dx___2 = as_factor(sa_dx___2)) |>
  mutate(sa_dx___3 = as_factor(sa_dx___3)) |>
  mutate(sa_dx___4 = as_factor(sa_dx___4)) |>
  mutate(sa_dx___5 = as_factor(sa_dx___5)) |>
  mutate(sa_dx___6 = as_factor(sa_dx___6)) |>
  mutate(sa_dx___7 = as_factor(sa_dx___7)) |>
```

```
  mutate(sa_dx___8 = as_factor(sa_dx___8)) |>
  mutate(sa_dx___9 = as_factor(sa_dx___9))


discharged_SUD <- factored_SUD |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_SUD <- factored_SUD |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_SUD <- factored_SUD |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))

joined_SUD <- baseline_SUD |>
  inner_join(discharged_SUD, by="record_id") |>
  inner_join(followup_SUD, by="record_id")
```

```
# Filtering for subjects that are in each time

discharged_social_support <- raw_social_support |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_social_support <- raw_social_support |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_social_support <- raw_social_support |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))

joined_social_support <- baseline_social_support |>
  inner_join(discharged_social_support, by="record_id") |>
  inner_join(followup_social_support, by="record_id")
```

```
# Filtering for subjects that are in each time

discharged_sub_history <- raw_sub_history |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_sub_history <- raw_sub_history |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_sub_history <- raw_sub_history |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))

joined_sub_history <- baseline_sub_history |>
  inner_join(discharged_sub_history, by="record_id") |>
```

```
  inner_join(followup_sub_history, by="record_id")
```

```
# Filtering for subjects that are in each time

discharged_aana_affiliation <- raw_aana_affiliation |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_aana_affiliation <- raw_aana_affiliation |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_aana_affiliation <- raw_aana_affiliation |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))

joined_aana_affiliation <- baseline_aana_affiliation |>
  inner_join(discharged_aana_affiliation, by="record_id") |>
  inner_join(followup_aana_affiliation, by="record_id")
```

```
# Filtering for subjects that are in each time

discharged_stressful_life <- raw_stressful_life |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_stressful_life <- raw_stressful_life |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_stressful_life <- raw_stressful_life |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))

joined_stressful_life <- baseline_stressful_life |>
  inner_join(discharged_stressful_life, by="record_id") |>
  inner_join(followup_stressful_life, by="record_id")
```

```
# Filtering for subjects that are in each time

discharged_spiritual <- raw_spiritual |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_spiritual <- raw_spiritual |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_spiritual <- raw_spiritual |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))

joined_spiritual <- baseline_spiritual |>
  inner_join(discharged_spiritual, by="record_id") |>
```

```
  inner_join(followup_spiritual, by="record_id")
```

```
# Filtering for subjects that are in each time

discharged_childhood <- raw_childhood |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_childhood <- raw_childhood |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_childhood <- raw_childhood |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))

joined_childhood <- baseline_childhood |>
  inner_join(discharged_childhood, by="record_id") |>
  inner_join(followup_childhood, by="record_id")
```

```
# Filtering for subjects that are in each time

discharged_life_quality <- raw_life_quality |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_life_quality <- raw_life_quality |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_life_quality <- raw_life_quality |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))

joined_life_quality <- baseline_life_quality |>
  inner_join(discharged_life_quality, by="record_id") |>
  inner_join(followup_life_quality, by="record_id")
```

```
# Filtering for subjects that are in each time

discharged_craving <- raw_craving |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_craving <- raw_craving |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_craving <- raw_craving |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))

joined_craving <- baseline_craving |>
  inner_join(discharged_craving, by="record_id") |>
```

```
  inner_join(followup_craving, by="record_id")
```

```
# Filtering for subjects that are in each time

discharged_change_impression <- raw_change_impression |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_change_impression <- raw_change_impression |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_change_impression <- raw_change_impression |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))

joined_change_impression <- baseline_change_impression |>
  inner_join(discharged_change_impression, by="record_id") |>
  inner_join(followup_change_impression, by="record_id")
```

```
# Filtering for subjects that are in each time

discharged_commitment <- raw_commitment |>
  filter(str_detect(redcap_event_name,'discharge')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".discharge")))

baseline_commitment <- raw_commitment |>
  filter(str_detect(redcap_event_name,'baseline')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".baseline")))

followup_commitment <- raw_commitment |>
  filter(str_detect(redcap_event_name,'followup')) |>
  rename_with(~ ifelse(.x == "record_id", .x, paste0(.x, ".followup")))

joined_commitment <- baseline_commitment |>
  inner_join(discharged_commitment, by="record_id") |>
  inner_join(followup_commitment, by="record_id")
```

Joining all

```
all_addiction_data <- inner_join(joined_demographics, joined_aana_affiliation, by="record_id") |>
  inner_join(joined_change_impression, by="record_id") |>
  inner_join(joined_commitment, by="record_id") |>
  inner_join(joined_childhood, by="record_id") |>
  inner_join(joined_craving, by="record_id") |>
  inner_join(joined_life_quality, by="record_id") |>
  inner_join(joined_social_support, by="record_id") |>
  inner_join(joined_spiritual, by="record_id") |>
  inner_join(joined_stressful_life, by="record_id") |>
  inner_join(joined_sub_history, by="record_id") |>
  inner_join(joined_SUD, by="record_id")
```

```r
saveRDS(all_addiction_data, "addiction.rds")
```