GMS6025C: Final Project 1 Part 2 Exploratory Data Analysis

Emely Gazarov, Ryo Iwata, Maria Ramirez

04 May, 2024

About the data

Name: Addiction Patient Assessment Measures

Description: Data are gathered from a residential treatment facility for substance use disorders. Data include item-level responses for all questionnaires listed below. Data are available at initiation of treatment, after one month of treatment, and at treatment discharge.

Overview of available datasets

Independent variables

Basic Demographics

- age
  - `age`
- gender
  - `gender`
- education
  - `education`
- number of days sober
  - `sober_days`
- number of days in treatment
  - `treatment_days`
- If a subject dropped out of the program or not
  - `dropout_yn`

Substance use disorders

- Flag of whether or not the subject is being treated for substance in name
- `SUD_is_Alchohol`
- `SUD_is_Opioid`
- `SUD_is_Cannabis`
- `SUD_is_depressants_anxiolytic`
- `SUD_is_Cocaine`
- `SUD_is_Other_stimulant`
- `SUD_is_Hallucinogen`
- `SUD_is_Nicotine`

- `SUD_is_Inhalant`
- `SUD_is_psychoactive`
- Number of substances being treated for
  - `SUD.sum`
    - Total number of all substances that subject is being treated for
  - `SUD.sum_legal`
    - Total number of legal substances that subject is being treated for. (SUD_is_Alchohol, SUD_is_Cannabis, SUD_is_Nicotine, SUD_is_Inhalant)
  - `SUD.sum_illegal`
    - Total number of legal substances that subject is being treated for. (SUD_is_Opioid, SUD_is_depressants_anxiolytic, SUD_is_Cocaine, SUD_is_Other_stimulant, SUD_is_psychoactive, SUD_is_Hallucinogen)
- Flag of whether or not the subject is being treated for legal or illegals substances
  - `SUD.uses_legal`
    - If subject is being treated for any of the following: SUD_is_Alchohol, SUD_is_Cannabis, SUD_is_Nicotine, SUD_is_Inhalant
  - `SUD.uses_illegal`
    - If subject is being treated for any of the following: SUD_is_Opioid, SUD_is_depressants_anxiolytic, SUD_is_Cocaine, SUD_is_Other_stimulant, SUD_is_psychoactive, SUD_is_Hallucinogen

Beliefs about AA/NA (TSPEQ)

- Categories based on number of AA/NA meetings attended throughout life
  - `aana_life`
- Categories based on number of AA/NA meetings attended last year
  - `aana_pastyear`
- Total score of questions gauging positive opinions of AA/NA
  - `aana_positive`
- Total score of questions gauging negative opinions of AA/NA
  - `aana_negative`

Childhood Experiences (ACE)

- Binary flags of whether the specified type of negative experience happened during childhood of subject
  - `childhood.verbal_abuse`
  - `childhood.physical_abuse`
  - `childhood.sexual_abuse`
  - `childhood.alone`
  - `childhood.neglected`
  - `childhood.divorced`
  - `childhood.parent_was_abused`
  - `childhood.other_was_addicted`
  - `childhood.other_was_stressed`
  - `childhood.other_was_prisoned`
- Sum of number of types of negative experiences that happened during subject's childhood
  - `childhood_sum`

Commitment to Sobriety (CSS-5)

- Total score for all Commitment to Sobriety (CSS-5) scores
  - Questions:
    1. Staying sober is the most important thing in my life.

2. I am totally committed to staying off alcohol/drugs.
3. I will do whatever it takes to recover from my addiction.
4. I never want to return to alcohol/drug use again.
5. I have had enough alcohol and drugs.

- Variables
  - `commit.BL` (baseline)
  - `commit.DC` (discharge)
  - `commit.FU` (followup)

•

Cravings

Please read each statement

and indicate how confident you are RIGHT NOW

that you would not choose to use a drug or drugs

if they were readily available to you TODAY.

1, Not at all confident | 2, Not Very confident | 3, Moderately confident |

4, Very confident | 5, Extremely confident

1. When I am in agony because of stopping or withdrawing from drug use.

2. When I have a headache.

3. When I am feeling depressed.

4. When I am on vacation and want to relax.

5. When I am concerned about someone.

6. When I am worried.

7. When I have the urge to use drugs to see what happens.

8. When I am being offered drugs in a social situation.

9. When I dream about using drugs.

10. When I want to test my will power over using drugs.

11. When I am feeling a physical need or craving for drugs.

12. When I am physically tired.

13. When I am experiencing some physical pain or injury.

14. When I feel like blowing up because of frustration.

15. When I see others using drugs at a bar or a party.

16. When I sense everything is going wrong for me.

17. When people I used to use drugs with encourage me to use drugs.

”

- Social support
  - family
  - friends
  - sig other
  - total
- Substance use history
  - Tried tobacco/alcohol
  - Age of first use
  - Regular use
  - Age of regular use
- Stressful life
  - happened
  - witnessed
  - learned about
  - exposed
  - total of all things
- Religion
  - Religious affiliation
  - Positive
  - Negative
- Life quality

Dependent variables

- impression change
- length of stay
- Life quality
- Commitment to change
- Cravings

```
cleaned_addiction_data = read_rds("./addiction.rds")
```

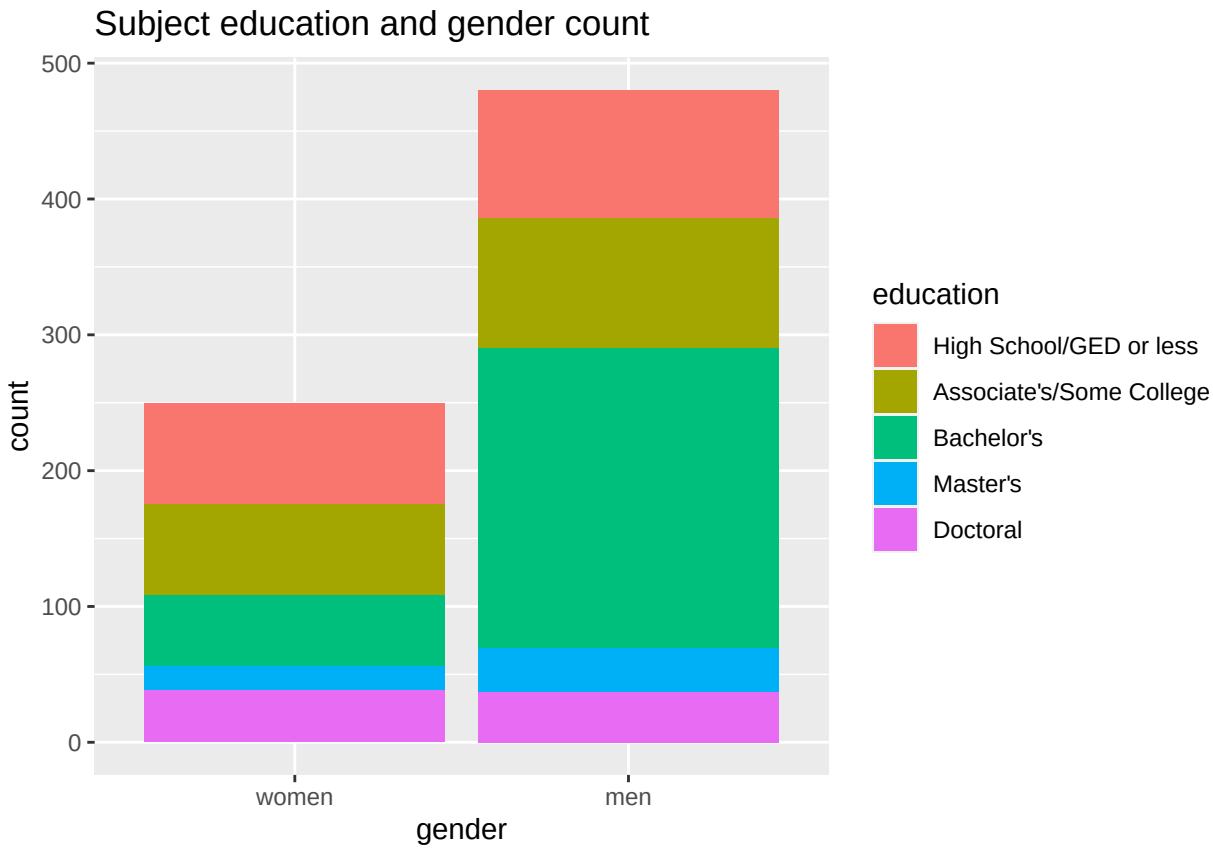
```
colnames(select(cleaned_addiction_data, starts_with("commit")))
```

```
## [1] "commit.BL" "commit.DC" "commit.FU"
```

Looking at the basic demographics

Education distribution

```
cleaned_addiction_data |> ggplot(aes(fill=education, x=gender)) +
  geom_bar() +
  labs(title = "Subject education and gender count")
```

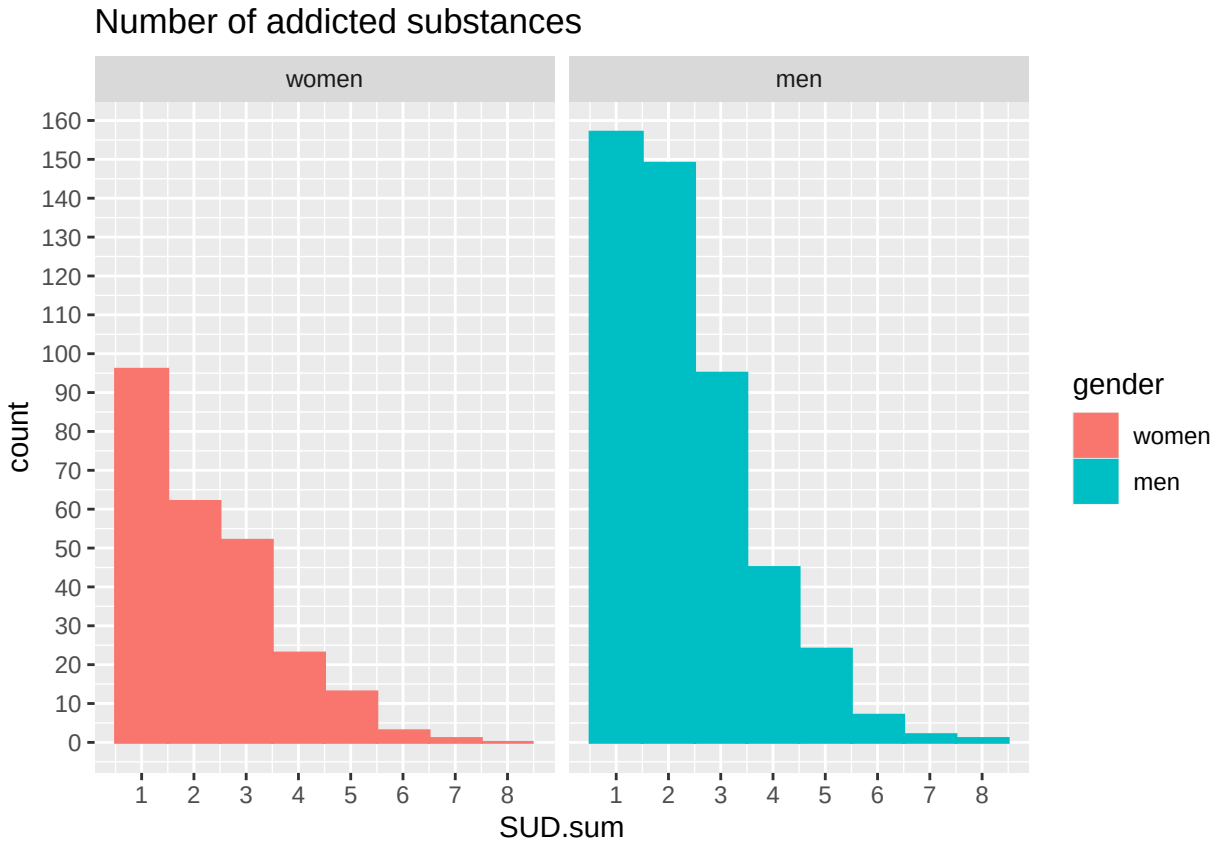


Age distribution

```
cleaned_addiction_data |>
  ggplot(aes(gender, age, color=gender)) +
  geom_violin() +
  geom_jitter(height = 0, width = 0.3) +
  stat_summary(fun = "mean",
               geom = "crossbar",
               color = "red") +
  # geom_boxplot(width=0.1) +
  labs(title = "Age distribution")
```

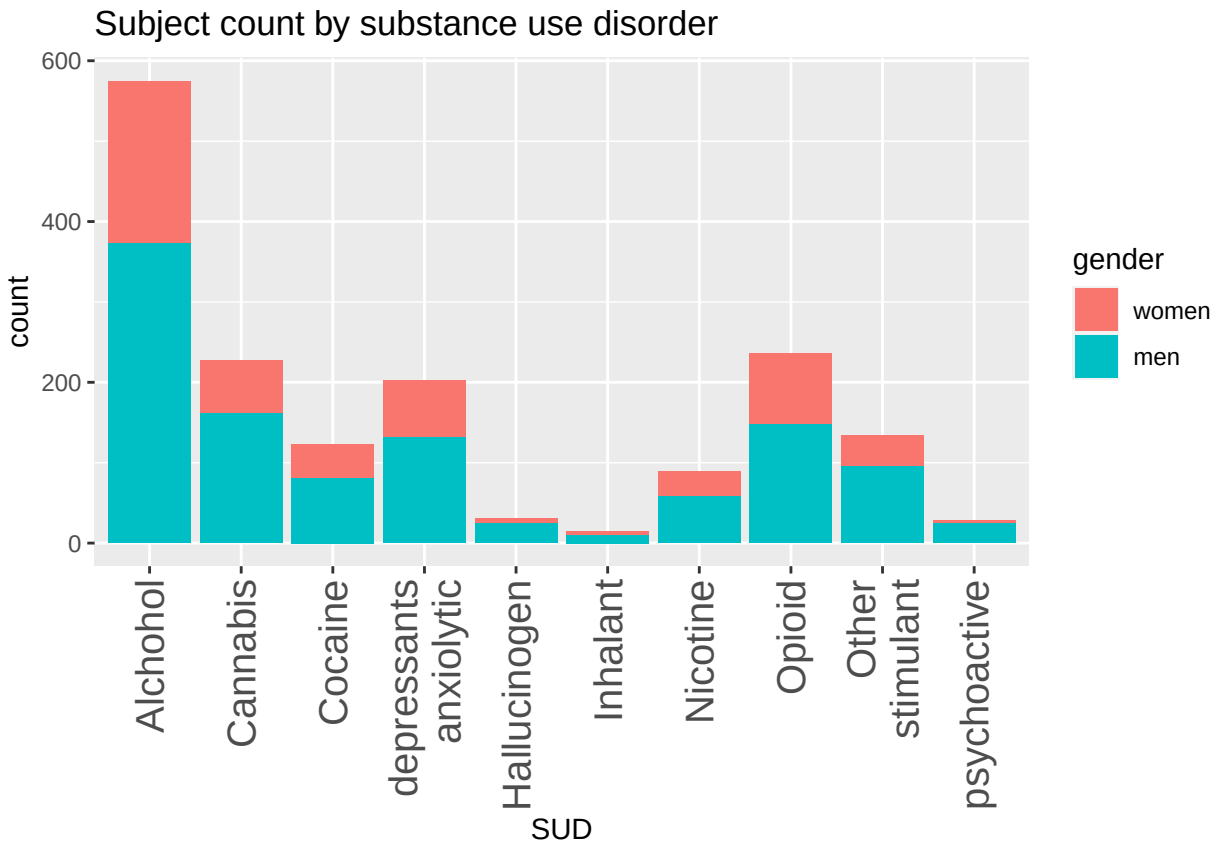
Distribution of number of addicted substances

```
cleaned_addiction_data |> ggplot(aes(SUD.sum, color=gender, fill=gender)) +
  geom_histogram(bins = 8) +
  facet_wrap(vars(gender)) +
  scale_x_continuous(n.breaks=10) +
  scale_y_continuous(n.breaks=20) +
  labs(title = "Number of addicted substances")
```



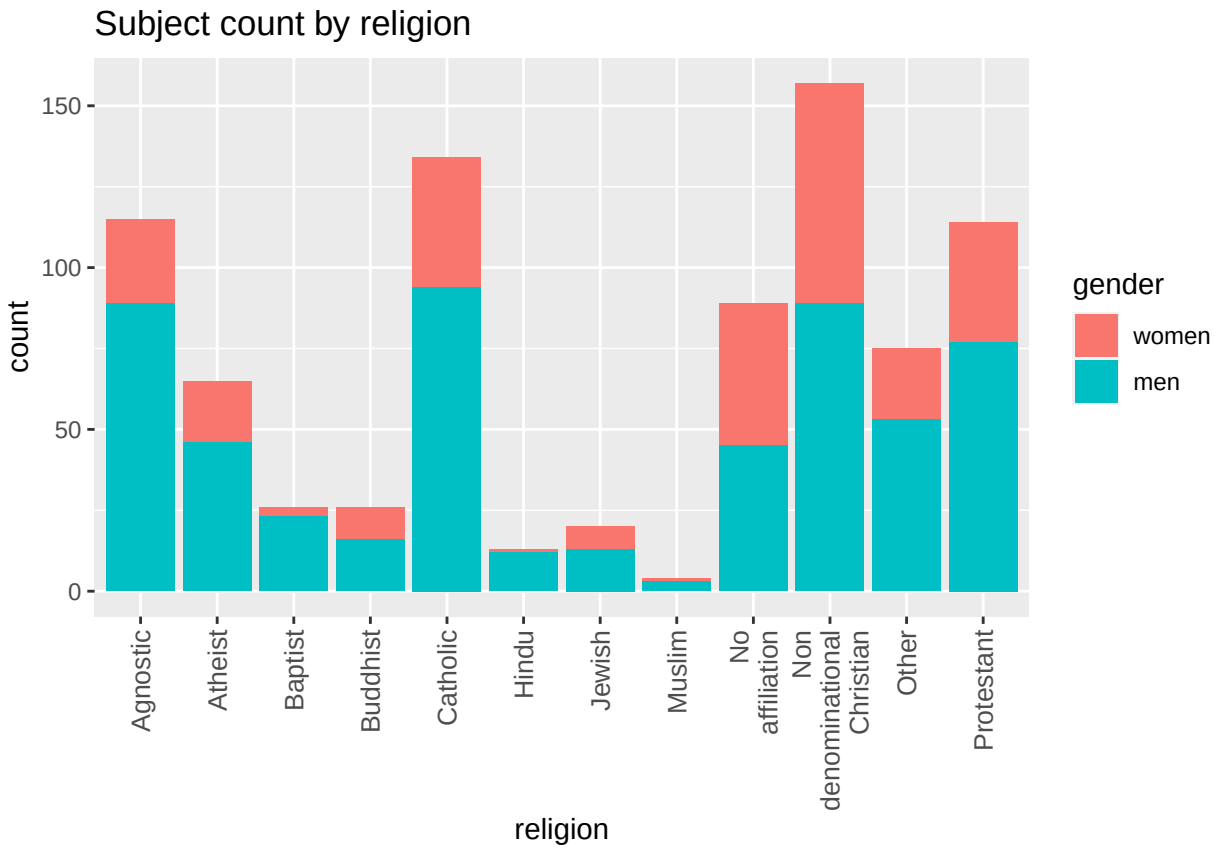
Number of subjects addicted to each substance

```
SUD_count_df <- cleaned_addiction_data |>
  pivot_longer(colnames(select(cleaned_addiction_data, starts_with("SUD_is"))),
               names_to = "SUD", values_to = "SUD_yn") |>
  filter(SUD_yn == "1") |>
  mutate(SUD = str_replace(SUD, "SUD_is_", "")) |>
  mutate(SUD = str_replace(SUD, "_", "\n"))


SUD_count_df |> ggplot(aes(fill=gender, x=SUD)) +
  geom_bar() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1, size=15)) +
  labs(title = "Subject count by substance use disorder")
```



Number of subjects practicing each religion

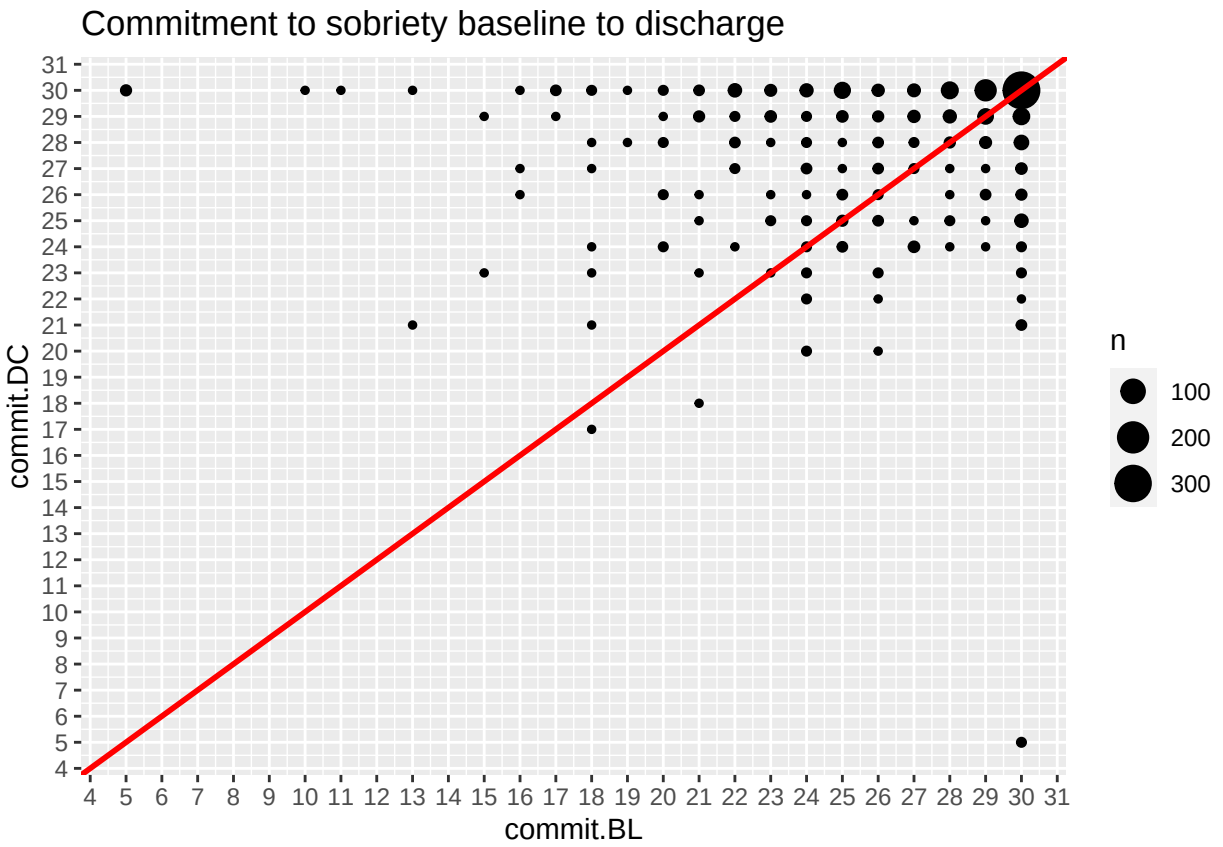
```
religion_count_df <- cleaned_addiction_data |>
  pivot_longer(colnames(select(cleaned_addiction_data, starts_with("rel.is_"))),
               names_to = "religion", values_to = "religion_yn") |>
  filter(religion_yn == "1") |>
  mutate(religion = str_replace(religion, "rel.is_", "")) |>
  mutate(religion = str_replace(religion, "_", "\n")) |>
  mutate(religion = str_replace(religion, "_", "\n"))

religion_count_df |> ggplot(aes(fill=gender, x=religion)) +
  geom_bar() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1, size=10)) +
  labs(title = "Subject count by religion")
```



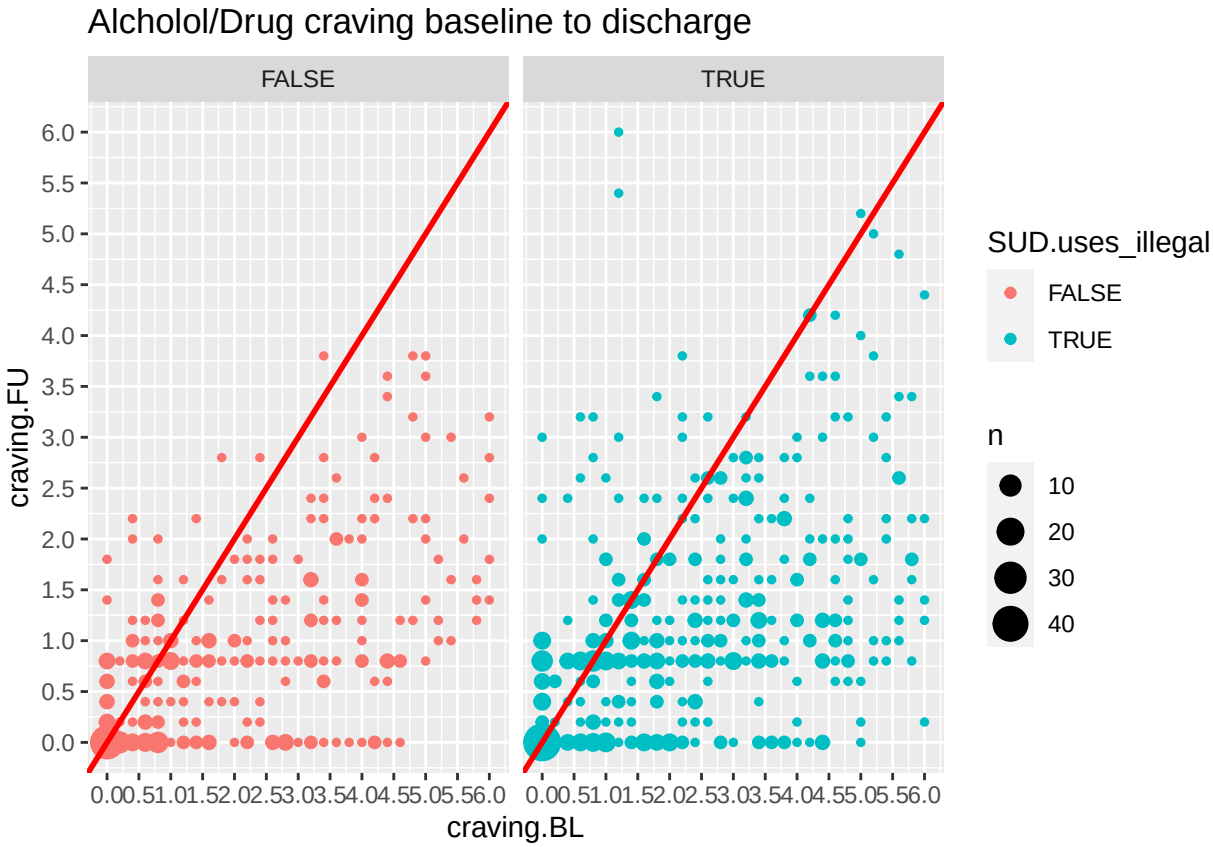
Looking at correlations

```
# TODO STARTING FROM HERE

cleaned_addiction_data |>
  ggplot(aes(commit.BL, commit.DC)) +
  geom_count() +
  scale_x_continuous(n.breaks=29) +
  scale_y_continuous(n.breaks=29) +
  geom_abline(slope=1, intercept = 0, color="red", size=1, alpha=3) +
  labs(title = "Commitment to sobriety baseline to discharge")
```

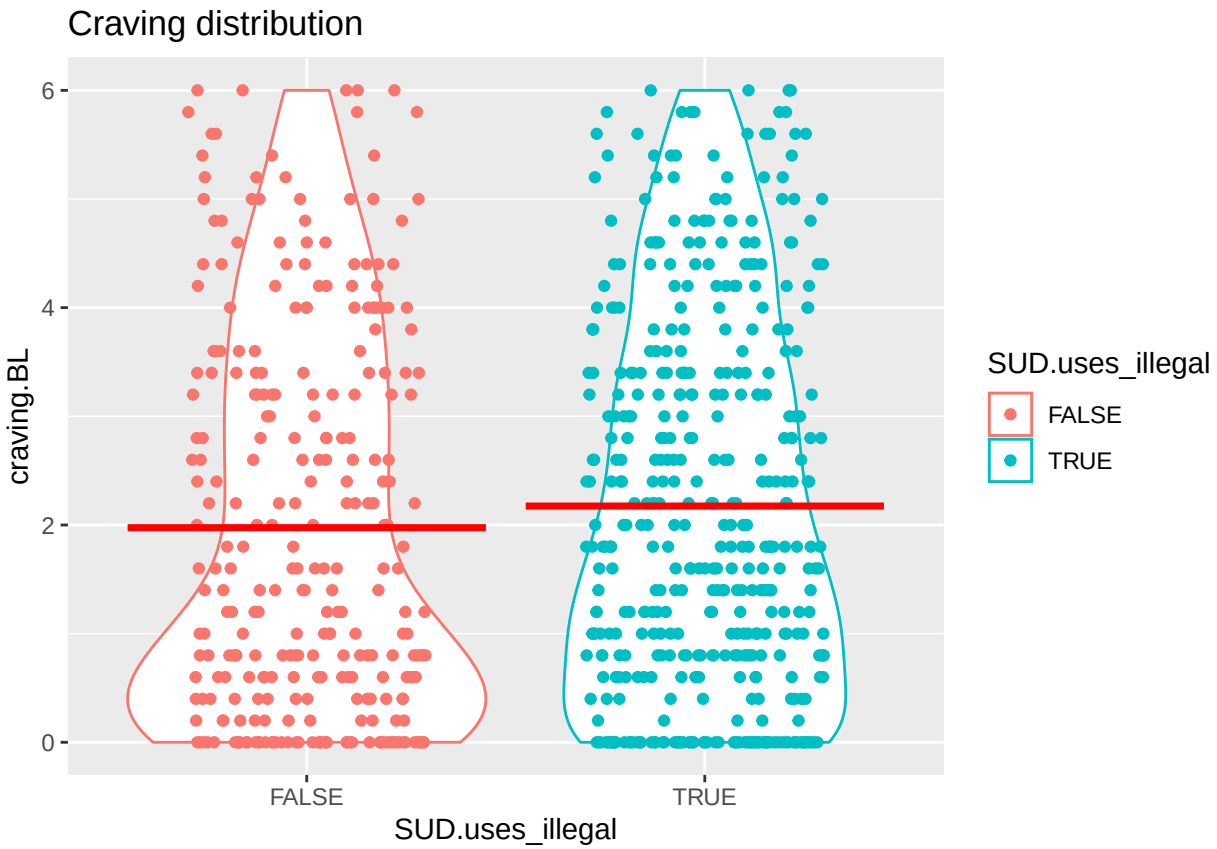
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



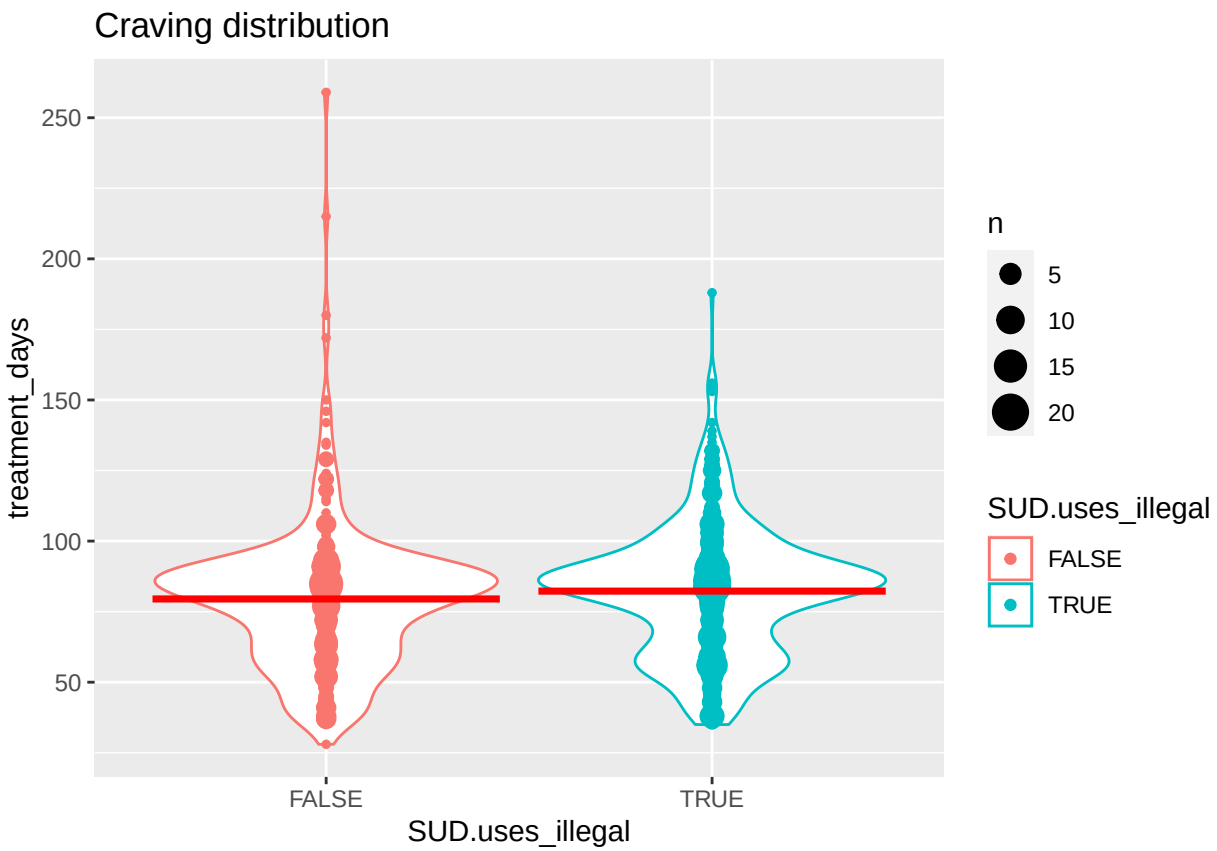

```
cleaned_addiction_data |>
  ggplot(aes(craving.BL, craving.FU, color=SUD.uses_illegal)) +
  geom_count() +
  scale_x_continuous(n.breaks=20) +
  scale_y_continuous(n.breaks=20) +
  geom_abline(slope=1, intercept = 0, color="red", size=1, alpha=3) +
  labs(title = "Alcholol/Drug craving baseline to discharge") +
  facet_wrap(vars(SUD.uses_illegal))
```




```
cleaned_addiction_data |>
  ggplot(aes(SUD.uses_illegal, craving.BL, color=SUD.uses_illegal)) +
  geom_violin() +
  geom_jitter(height = 0, width = 0.3) +
  # geom_count() +
  stat_summary(fun = "mean",
               geom = "crossbar",
               color = "red") +
  # geom_boxplot(width=0.1) +
  labs(title = "Craving distribution")
```

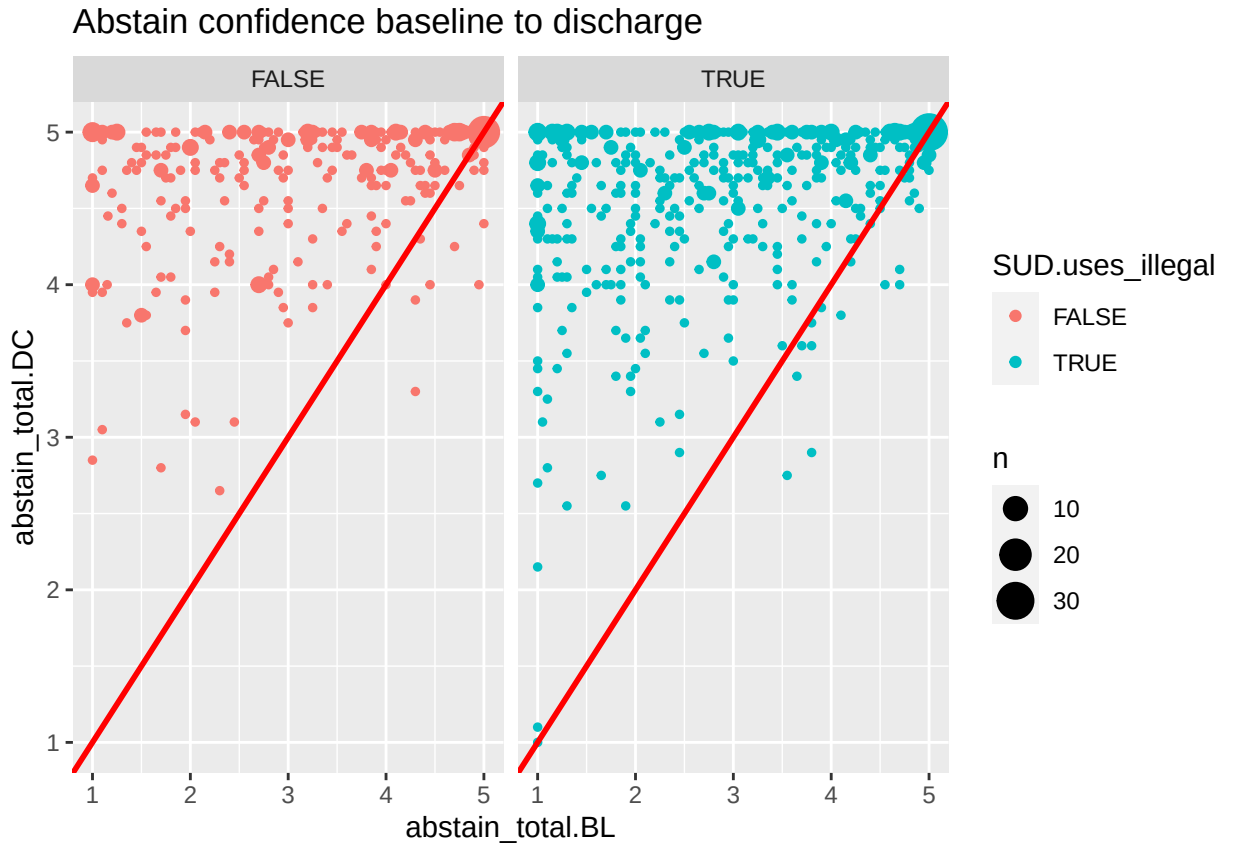



```
cleaned_addiction_data |>
  ggplot(aes(SUD.uses_illegal, treatment_days, color=SUD.uses_illegal)) +
  geom_violin() +
  # geom_jitter(height = 0, width = 0.1) +
  geom_count() +
  stat_summary(fun = "mean",
               geom = "crossbar",
               color = "red") +
  # geom_boxplot(width=0.1) +
  labs(title = "Craving distribution")
```

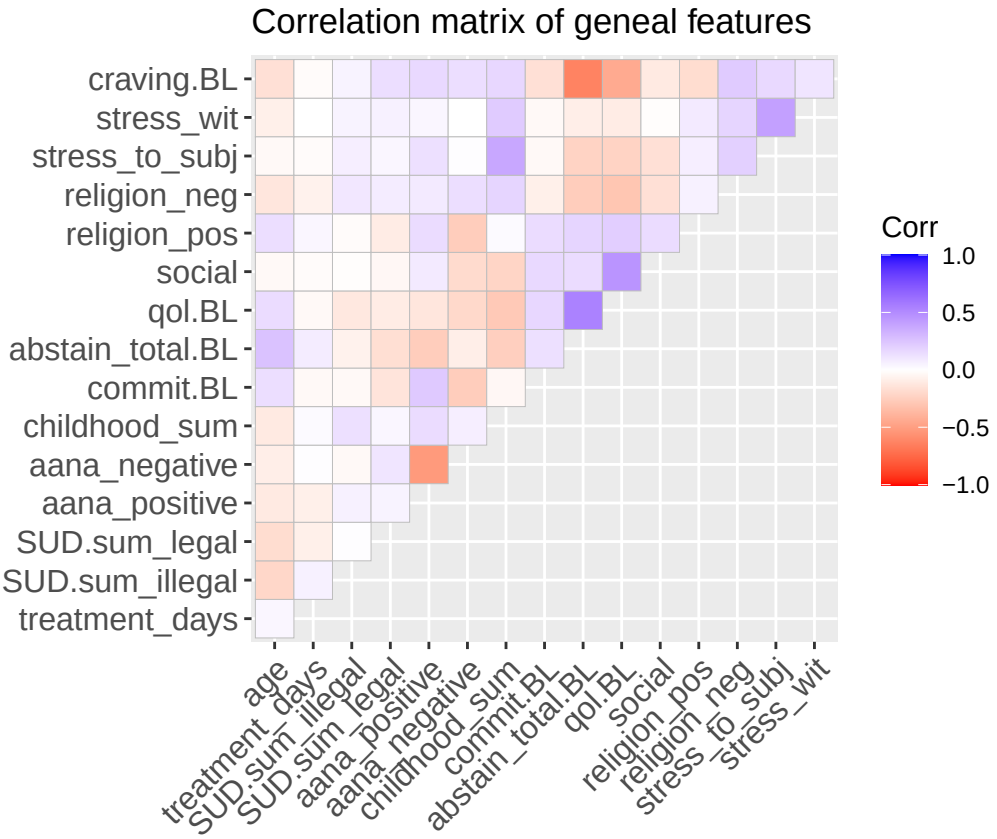



```
  # facet_wrap(vars(SUD.uses_illegal))
```

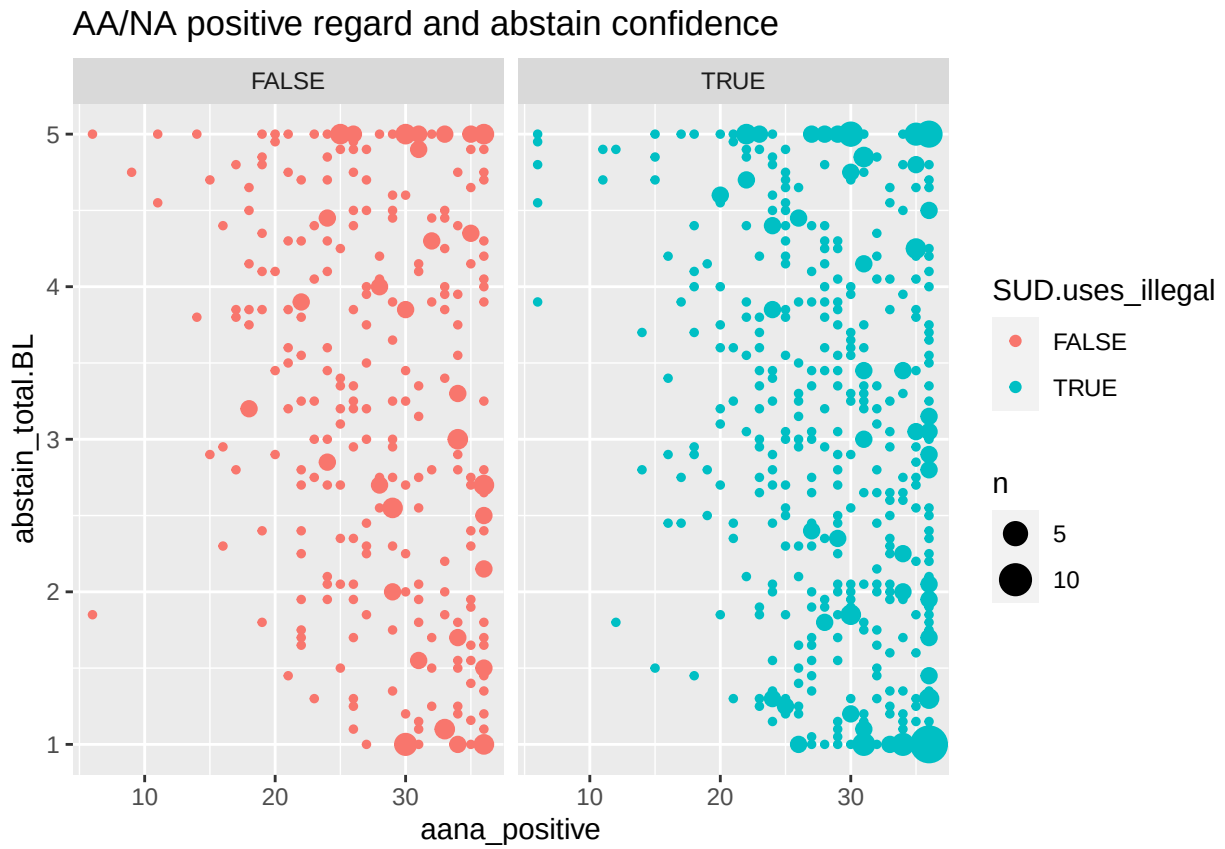
```
cleaned_addiction_data |>
  ggplot(aes(abstain_total.BL, abstain_total.DC, color=SUD.uses_illegal)) +
  geom_count() +
  # scale_x_continuous(n.breaks=20) +
  # scale_y_continuous(n.breaks=20) +
  geom_abline(slope=1, intercept = 0, color="red", size=1, alpha=3) +
  labs(title = "Abstain confidence baseline to discharge") +
  facet_wrap(vars(SUD.uses_illegal))
```



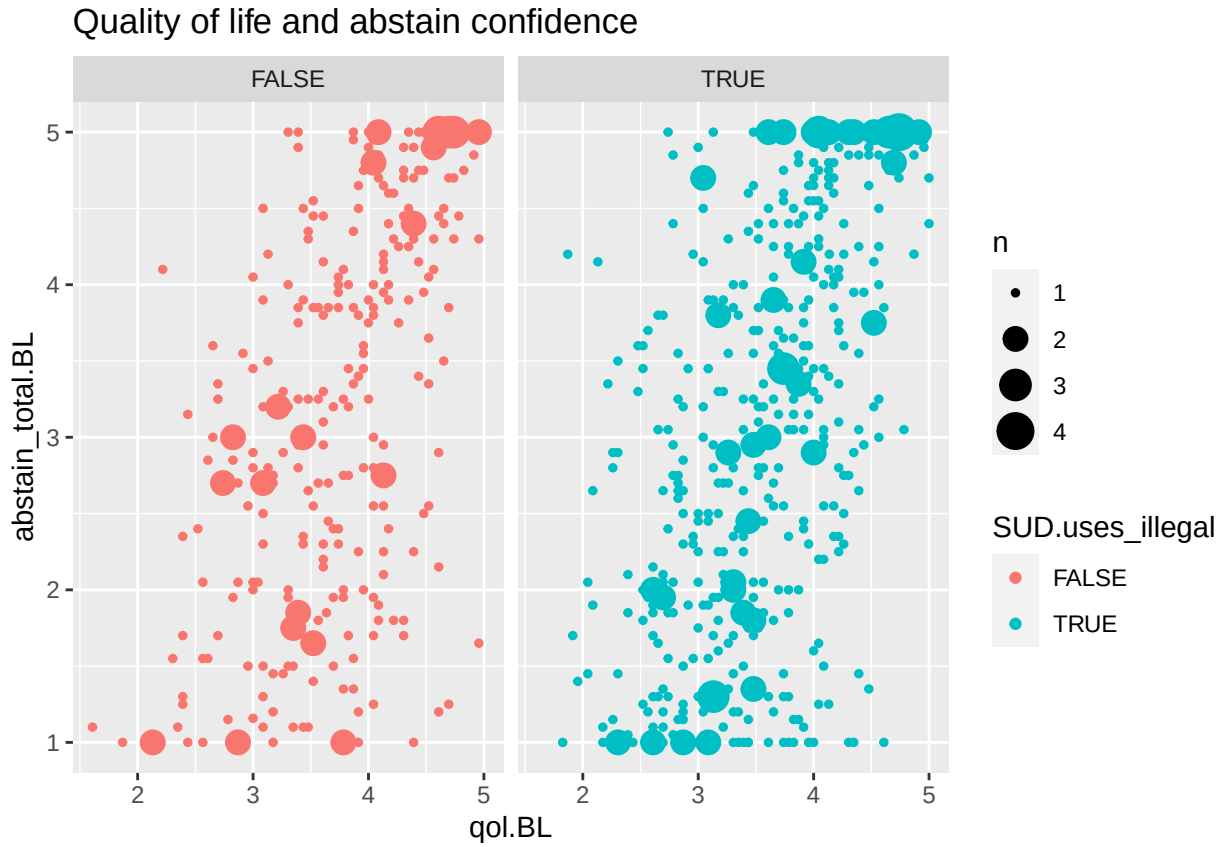

```
cleaned_addiction_data |>
  # select(ends_with("BL")) |>
  select(age, treatment_days, SUD.sum_illegal, SUD.sum_legal, aana_positive, aana_negative, childhood_su

  cor() |>
  ggcorrplot(colors = c("red", "white", "blue"),
             type = "upper",
             ggtheme = ggplot2::theme_gray,
             # hc.order = TRUE
             ) +

  labs(title = "Correlation matrix of geneal features")
```



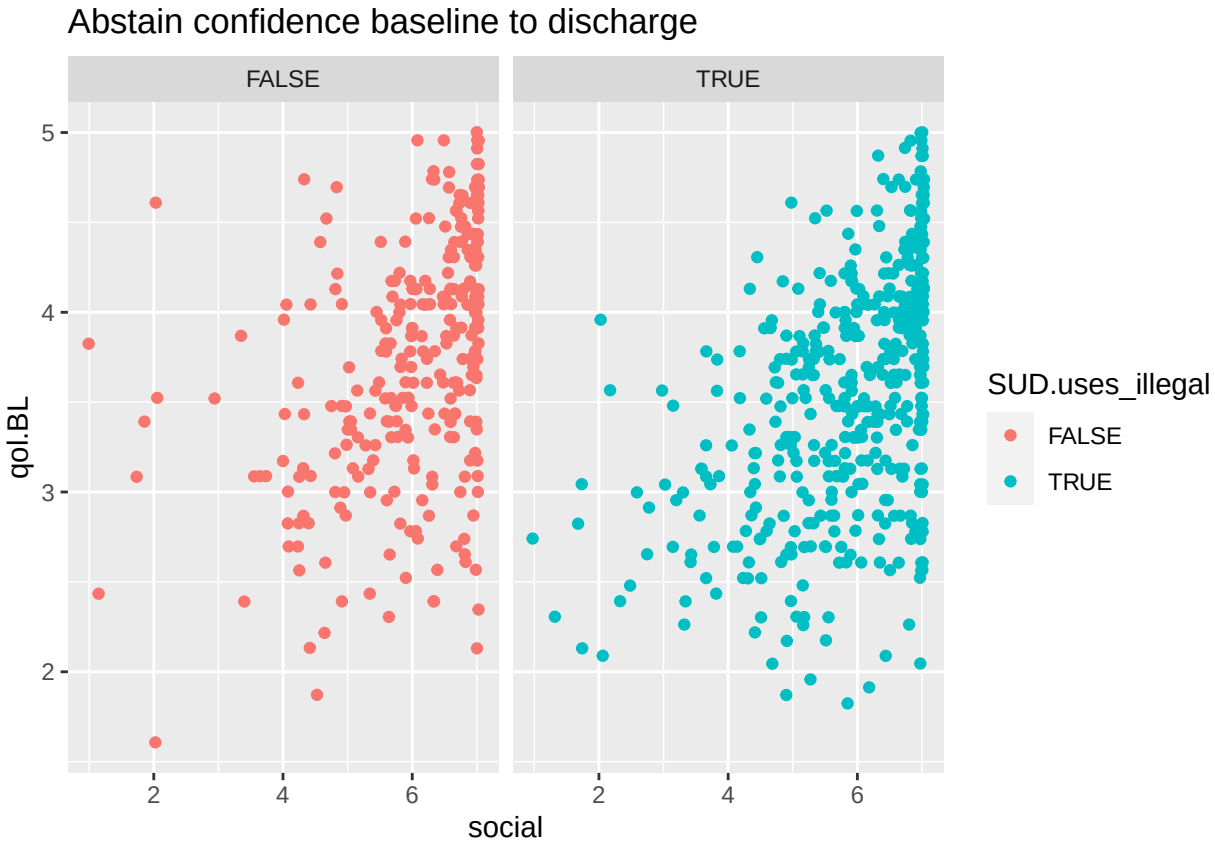

```
cleaned_addiction_data |>
  ggplot(aes(aana_positive, abstain_total.BL, color=SUD.uses_illegal)) +
  geom_count() +
  # scale_x_continuous(n.breaks=20) +
  # scale_y_continuous(n.breaks=20) +
  labs(title = "AA/NA positive regard and abstain confidence") +
  facet_wrap(vars(SUD.uses_illegal))
```



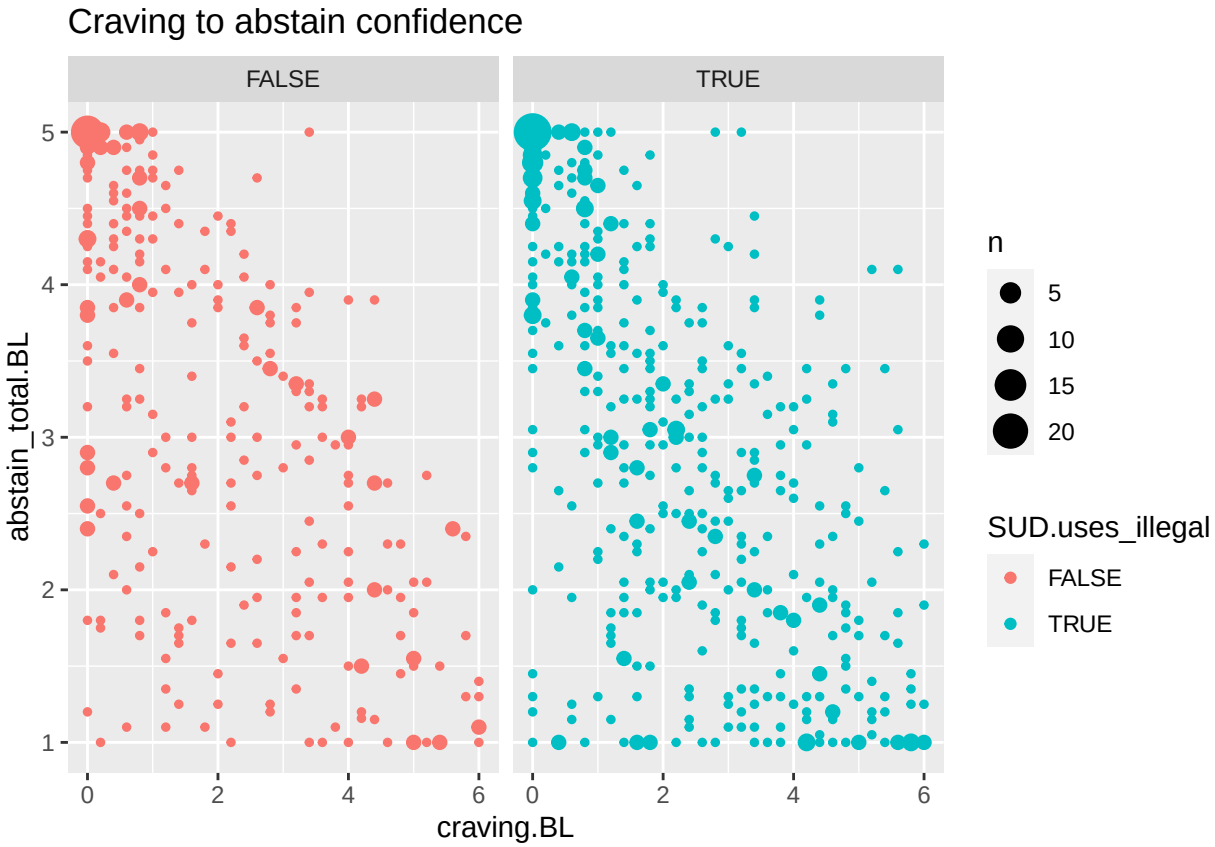

```
cleaned_addiction_data |>
  ggplot(aes(qol.BL, abstain_total.BL, color=SUD.uses_illegal)) +
  geom_count() +
  # scale_x_continuous(n.breaks=20) +
  # scale_y_continuous(n.breaks=20) +
  labs(title = "Quality of life and abstain confidence") +
  facet_wrap(vars(SUD.uses_illegal))
```




```
cleaned_addiction_data |>
  ggplot(aes(social, qol.BL, color=SUD.uses_illegal)) +
  # geom_count() +
  geom_jitter() +
  # scale_x_continuous(n.breaks=20) +
  # scale_y_continuous(n.breaks=20) +
  labs(title = "Abstain confidence baseline to discharge") +
  facet_wrap(vars(SUD.uses_illegal))
```




```
cleaned_addiction_data |>
  ggplot(aes(craving.BL, abstain_total.BL, color=SUD.uses_illegal)) +
  geom_count() +
  # scale_x_continuous(n.breaks=20) +
  # scale_y_continuous(n.breaks=20) +
  labs(title = "Craving to abstain confidence") +
  facet_wrap(vars(SUD.uses_illegal))
```


Craving to abstain confidence
FALSE
TRUE
abstain_total.BL
craving.BL
1
2
3
4
5
0
2
4
6
n
5
10
15
20
SUD.uses_illegal
FALSE
TRUE