

GMS6025C: Final Project 1 Part 2 Exploratory Data Analysis

Emely Gazarov, Ryo Iwata, Maria Ramirez

05 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

- Variables that end in .BL, .FU, .DC are indicating that the score was collected at baseline before treatment, 30 days following up after baseline, and during discharge respectively.

Independent variables

Basic Demographics

- age
 - age
- gender
 - gender
- education
 - education
- number of days sober
 - sober_days
- If a patient dropped out of the program or not
 - dropout_yn

Substance use disorders

- Flag of whether or not the patient is being treated for substance in name
 - SUD_is_Alcohol
 - 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_psychosactive
- Number of substances being treated for
 - Total number of all substances that patient is being treated for

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

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 patient
 - 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 patient's childhood
 - childhood_sum

Alcohol/Drug Abstinence Experiences (DAASE)

- Total score of questions that assesses patient's self-confidence in their ability to remain abstinent from alcohol in various situations.
 - Negative Affect
 - * abstain_neg.BL
 - * abstain_neg.FU
 - * abstain_neg.DC
 - Social/Positive
 - * abstain_pos.BL
 - * abstain_pos.FU

- * abstain_pos.DC
- Physical ailments
 - * abstain_phy.BL
 - * abstain_phy.FU
 - * abstain_phy.DC
- Situations associated with cravings
 - * abstain_crv.BL
 - * abstain_crv.FU
 - * abstain_crv.DC
- All previously mentioned conditions
 - * abstain_total.BL
 - * abstain_total.FU
 - * abstain_total.DC

Social Support (MSPSS)

- Total score for questions that measure the perceived level of social support the respondent receives from each of the three subcategories:
 - family
 - * social.family
 - friends
 - * social.friends
 - significant other
 - * social.sig_other
 - Sum of family, friends, and significant other
 - * social

Spiritual Experiences (Brief R-COPE)

- Total score of questions asking the frequency of spiritual coping experience, ranging from 0) Not at all to 3) Nearly every day.
 - Positive coping
 - * religion_pos
 - Negative coping
 - * religion_neg
- Religion
 - Binary flag of if a patient practices religion that is in the name of the variable
 - * rel.is_Atheist
 - * rel.is_Agnostic
 - * rel.is_Protestant
 - * rel.is_Catholic
 - * rel.is_Muslim
 - * rel.is_Jewish
 - * rel.is_Hindu
 - * rel.is_Buddhist
 - * rel.is_Baptist
 - * rel.is_No_affiliation
 - * rel.is_Non_denominational_Christian
 - * rel.is_Other
 - Binary flag of if patient is religious or not
 - * rel.is_religious
 - * rel.is_not_religious

Stressful Life Experiences (LEC-5)

- Binary flags of whether a listed stressful life event was either 1) experienced the life event personally, 2) witnessed it happen to someone else, 3) learned about it happening from a close friend or family member, 4) experienced the event due to their job.
 - Natural disaster
 - * `stress.natural_disaster.to_subj`
 - * `stress.natural_disaster.witnessed`
 - * `stress.natural_disaster.learned`
 - * `stress.natural_disaster.exposed`
 - Fire or explosion
 - * `stress.fire.to_subj`
 - * `stress.fire.witnessed`
 - * `stress.fire.learned`
 - * `stress.fire.exposed`
 - Transportation accident
 - * `stress.transportation_accident.to_subj`
 - * `stress.transportation_accident.witnessed`
 - * `stress.transportation_accident.learned`
 - * `stress.transportation_accident.exposed`
 - Non-transportation accident
 - * `stress.serious_accident.to_subj`
 - * `stress.serious_accident.witnessed`
 - * `stress.serious_accident.learned`
 - * `stress.serious_accident.exposed`
 - Exposure to toxic substance
 - * `stress.toxic.to_subj`
 - * `stress.toxic.witnessed`
 - * `stress.toxic.learned`
 - * `stress.toxic.exposed`
 - Physically assaulted
 - * `stress.physical_assault.to_subj`
 - * `stress.physical_assault.witnessed`
 - * `stress.physical_assault.learned`
 - * `stress.physical_assault.exposed`
 - Assaulted with a weapon
 - * `stress.weapon_assault.to_subj`
 - * `stress.weapon_assault.witnessed`
 - * `stress.weapon_assault.learned`
 - * `stress.weapon_assault.exposed`
 - Sexual assault
 - * `stress.sexual_assault.to_subj`
 - * `stress.sexual_assault.witnessed`
 - * `stress.sexual_assault.learned`
 - * `stress.sexual_assault.exposed`
 - Unwanted or uncomfortable sexual experience.
 - * `stress.unwanted_sexual.to_subj`
 - * `stress.unwanted_sexual.witnessed`
 - * `stress.unwanted_sexual.learned`
 - * `stress.unwanted_sexual.exposed`
 - Combat or exposure to war-zone
 - * `stress.combat.to_subj`
 - * `stress.combat.witnessed`
 - * `stress.combat.learned`

- * stress.combat.exposed
- Captivity
 - * stress.captivity.to_subj
 - * stress.captivity.witnessed
 - * stress.captivity.learned
 - * stress.captivity.exposed
- Life-threatening illness or injury.
 - * stress.illness.to_subj
 - * stress.illness.witnessed
 - * stress.illness.learned
 - * stress.illness.exposed
- Severe human suffering.
 - * stress.severe_suffering.to_subj
 - * stress.severe_suffering.witnessed
 - * stress.severe_suffering.learned
 - * stress.severe_suffering.exposed
- Sudden violent death (for example, homicide, suicide).
 - * stress.sudden_violent_death.to_subj
 - * stress.sudden_violent_death.witnessed
 - * stress.sudden_violent_death.learned
 - * stress.sudden_violent_death.exposed
- Sudden accidental death.
 - * stress.sudden_accidental_death.to_subj
 - * stress.sudden_accidental_death.witnessed
 - * stress.sudden_accidental_death.learned
 - * stress.sudden_accidental_death.exposed
- Serious injury, harm, or death you caused to someone else.
 - * stress.harm_to_others.to_subj
 - * stress.harm_to_others.witnessed
 - * stress.harm_to_others.learned
 - * stress.harm_to_others.exposed
- Any other very stressful event or experience.
 - * stress.other.to_subj
 - * stress.other.witnessed
 - * stress.other.learned
 - * stress.other.exposed
- Total score of number of categories that happened to the patient personally and/or witnessed happen to someone else
 - * stress_to_subj
 - * stress_to_subj_and_wit
 - * stress_wit

Substance use history

- Binary flag of whether or not patient has used the following:
 - Tobacco
 - * history.tobacco
 - Alcohol
 - * history.alcohol
 - Other
 - * history.other

Dependent variables

Length of stay in treatment

- number of days in treatment
 - `treatment_days`

Alcohol/Drug Craving (PACS)

- Total score that assesses the alcohol/drug cravings the patient has experienced within the past week based on a variety of craving characteristics.
 - `craving.BL`
 - `craving.FU`
 - `craving.DC`

Quality of Life (WHOQOL-BREF)

- Total scores of assessment that measures different aspects of quality of life
 - general
 - * `qol_general_health.BL`
 - * `qol_general_health.FU`
 - * `qol_general_health.DC`
 - physical
 - * `qol_physical.BL`
 - * `qol_physical.FU`
 - * `qol_physical.DC`
 - psychological health
 - * `qol_psych.BL`
 - * `qol_psych.FU`
 - * `qol_psych.DC`
 - social relationships
 - * `qol_social.BL`
 - * `qol_social.FU`
 - * `qol_social.DC`
 - environment
 - * `qol_env.BL`
 - * `qol_env.FU`
 - * `qol_env.DC`

Commitment to Sobriety (CSS-5)

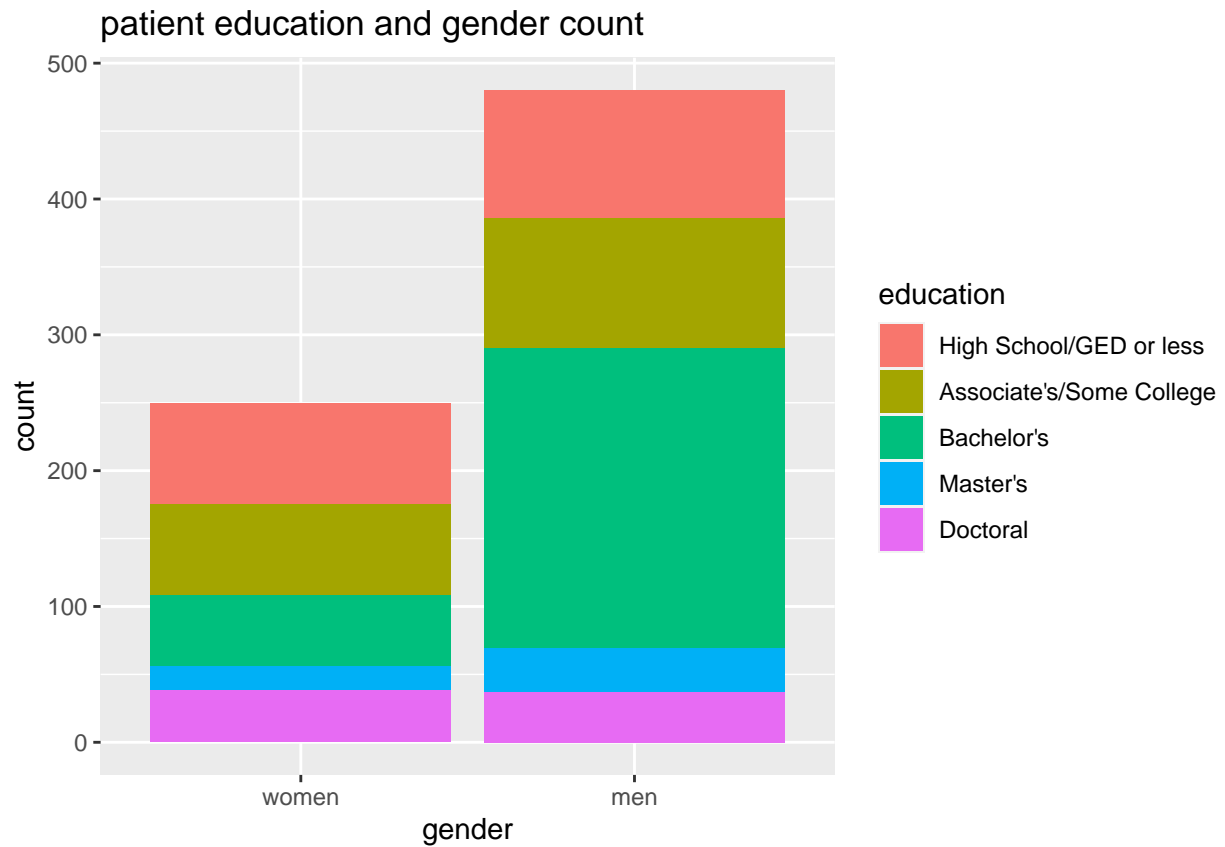
- Total score of questions gauging patient's commitment to abstinence from alcohol and drug use
 - `commit.BL`
 - `commit.DC`
 - `commit.FU`

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

Looking at the basic demographics

Education distribution

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



Age distribution

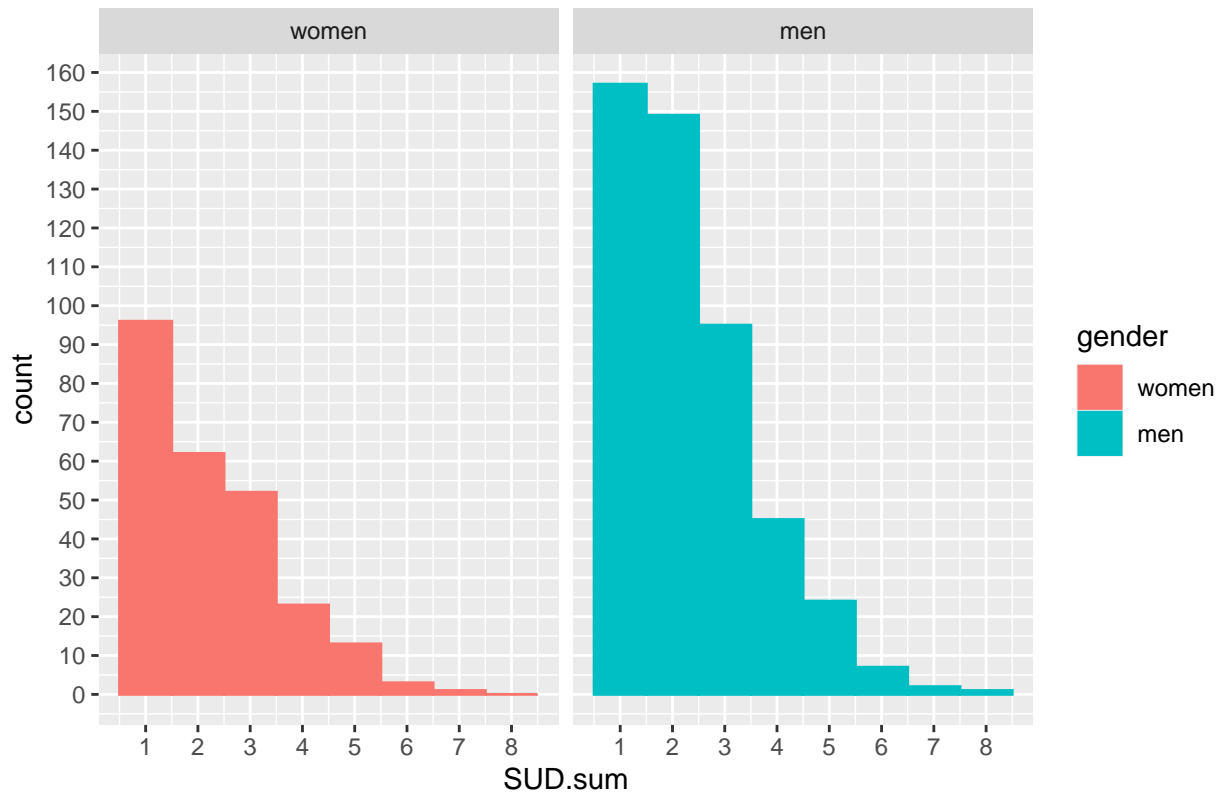
```
cleaned_addiction_data |>
  ggplot(aes(gender, age, color=gender)) +
  geom_violin() +
  geom_jitter(height = 0, width = 0.3) +
  stat_summary(fun = "mean",
    geom = "crossbar",
    color = "red") +
  # geom_boxplot(width=0.1) +
  labs(title = "Age distribution")
```



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")
```

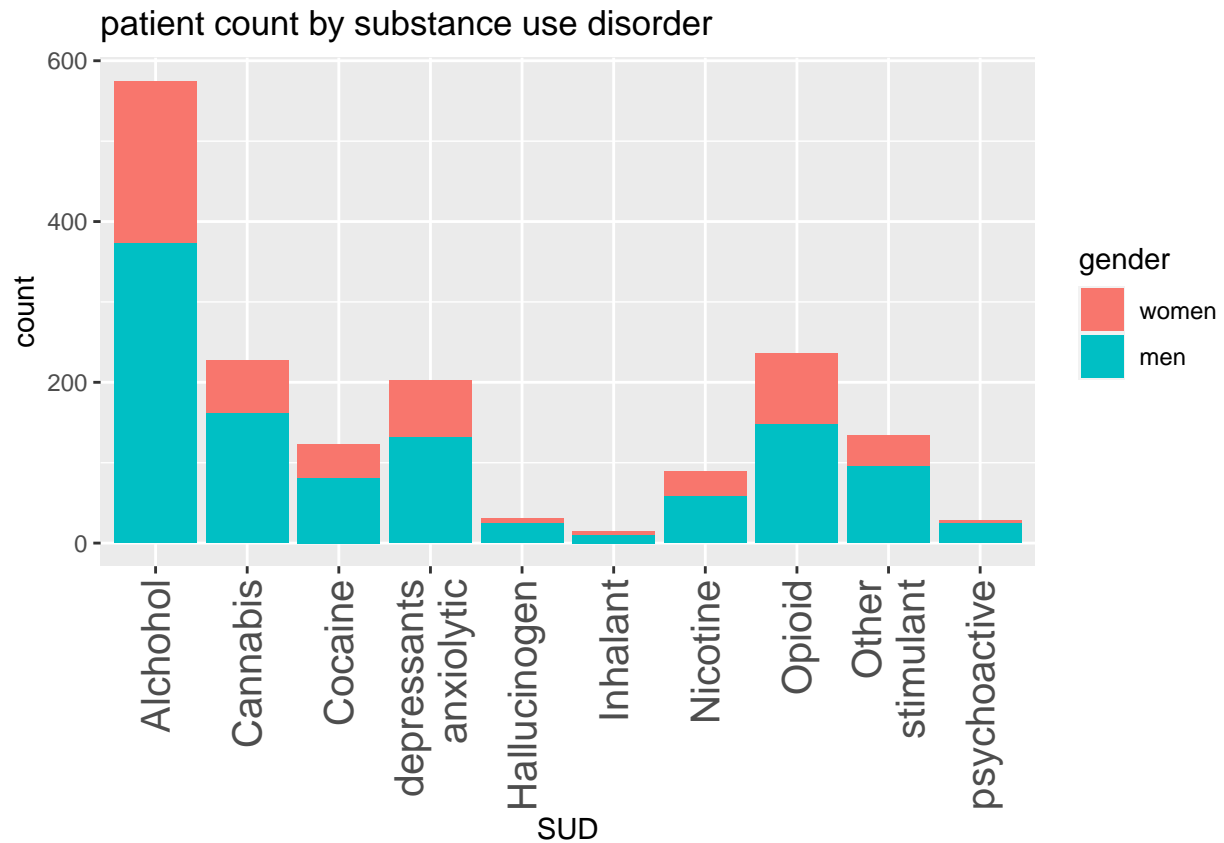

Number of addicted substances



Number of patients 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"))

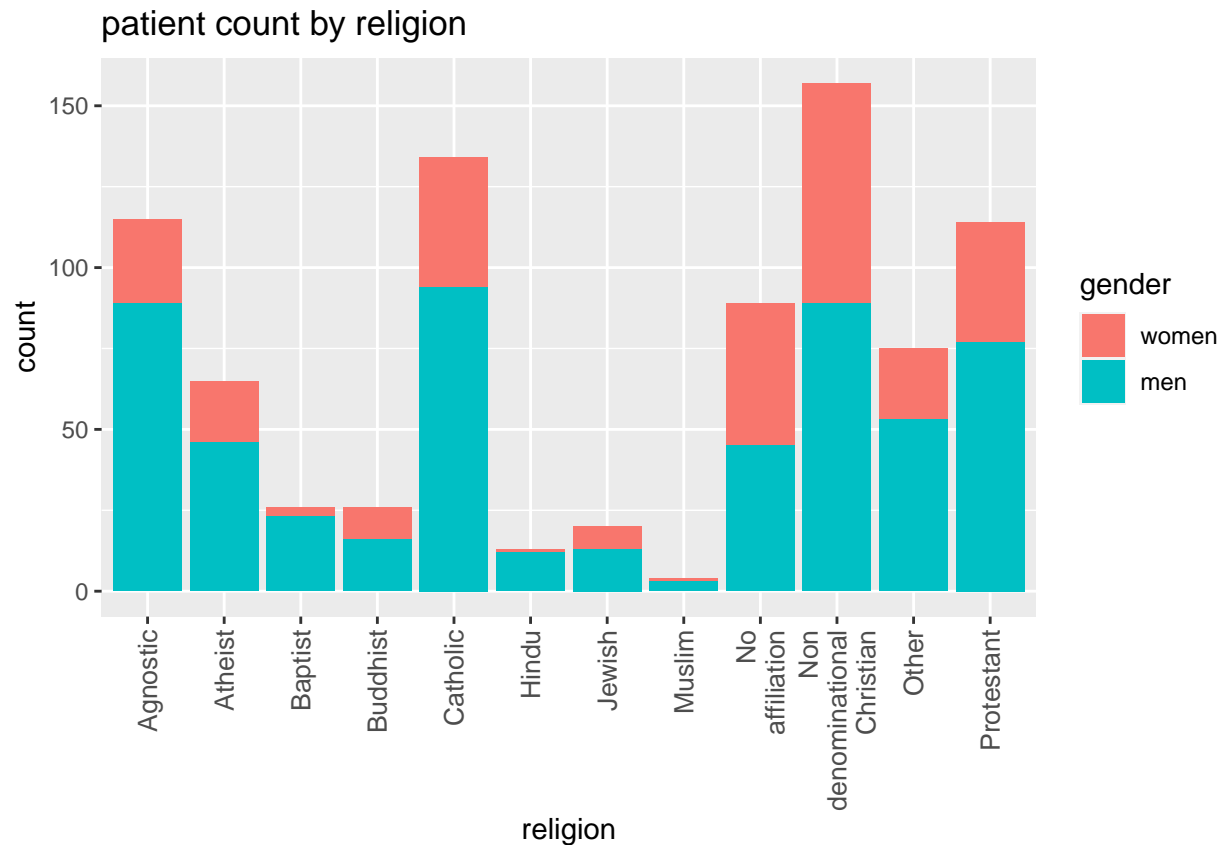
SUD_count_df |> ggplot(aes(fill=gender, x=SUD)) +
  geom_bar() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1, size=15)) +
  labs(title = "patient count by substance use disorder")
```



Number of patients 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 = "patient count by religion")
```



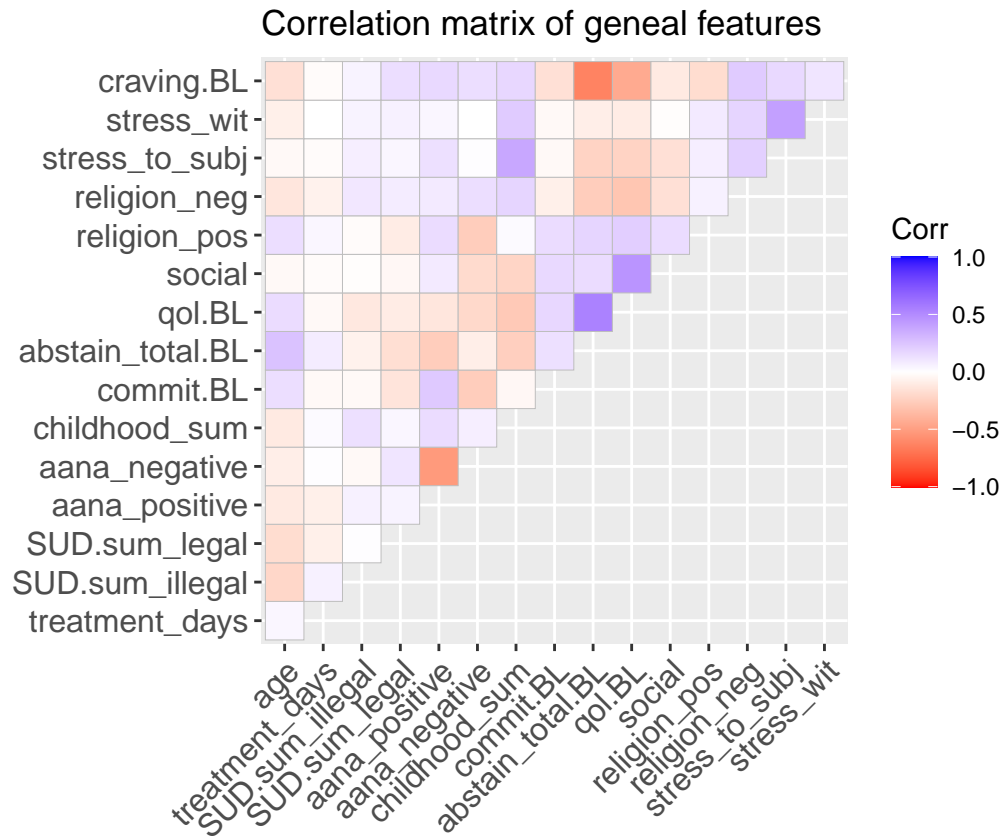
Looking at correlations

Correlations across continous variables

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

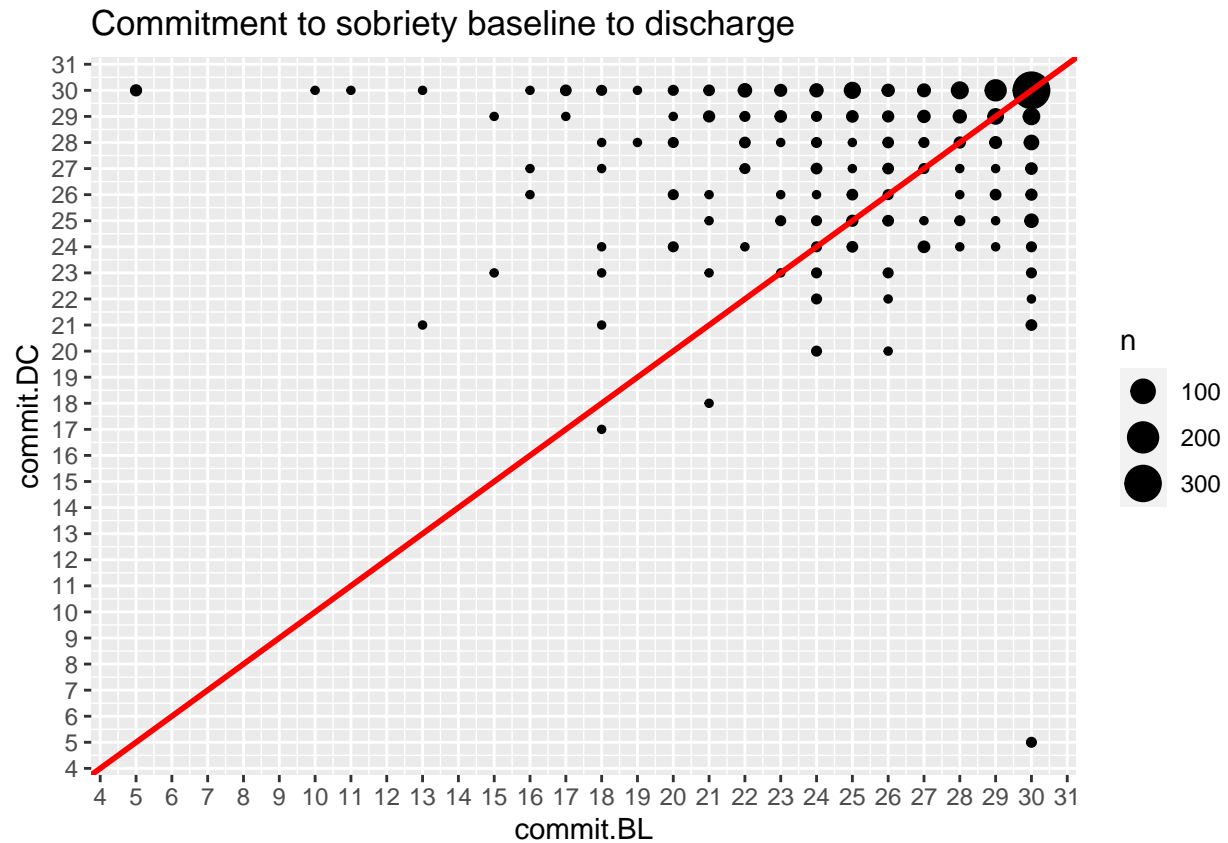
cor() |>
ggcorrplot(colors = c("red", "white", "blue"),
  type = "upper",
  ggtheme = ggplot2::theme_gray,
  # hc.order = TRUE
) +

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



Looking at relationship between commitment to sobriety at baseline to discharge

```
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", linewidth=1, alpha=3) +
  labs(title = "Commitment to sobriety baseline to discharge")
```



Looking at relationship between commitment to sobriety at baseline to discharge

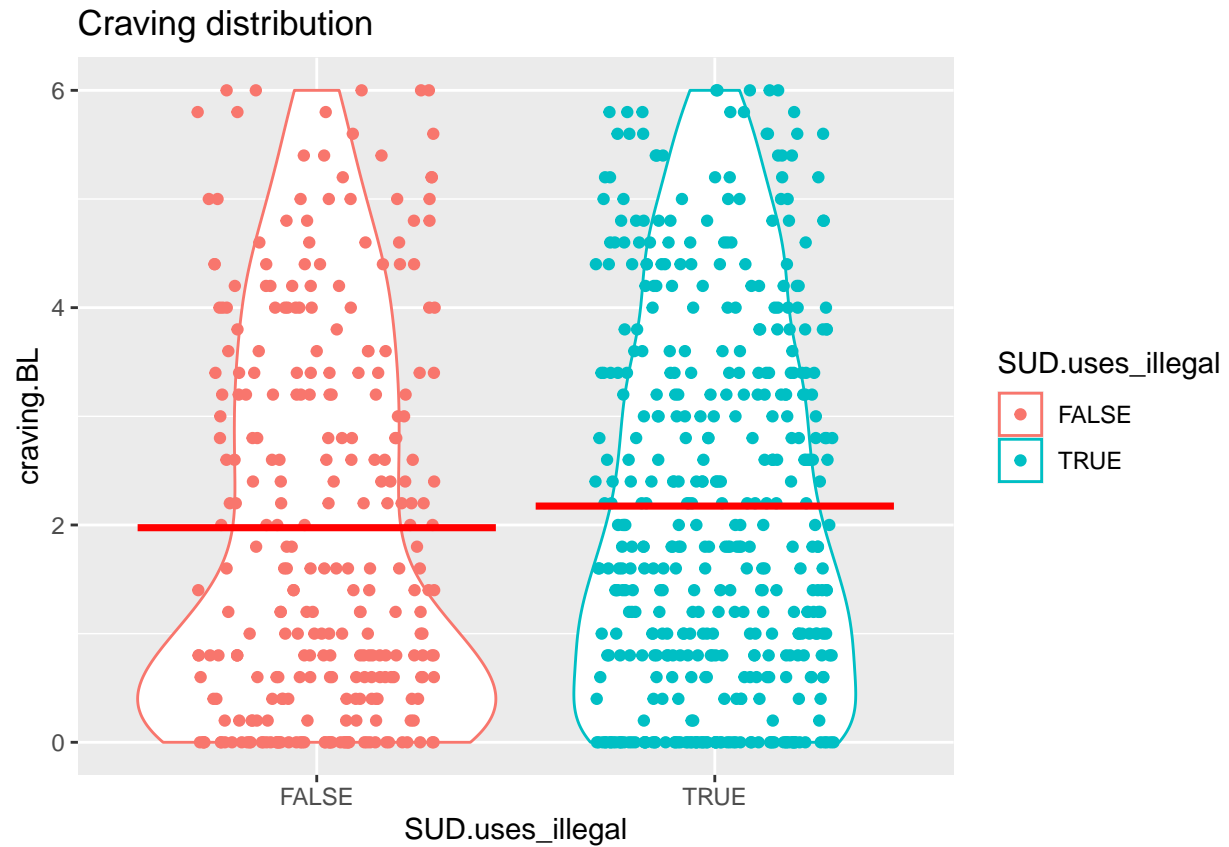
```
cleaned_addiction_data |>
  ggplot(aes(craving.BL, craving.FU, color=SUD.uses_illegal)) +
  geom_count() +
  scale_x_continuous(n.breaks=20) +
  scale_y_continuous(n.breaks=20) +
  geom_abline(slope=1, intercept = 0, color="red", size=1, alpha=3) +
  labs(title = "Alcohol/Drug craving baseline to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

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.

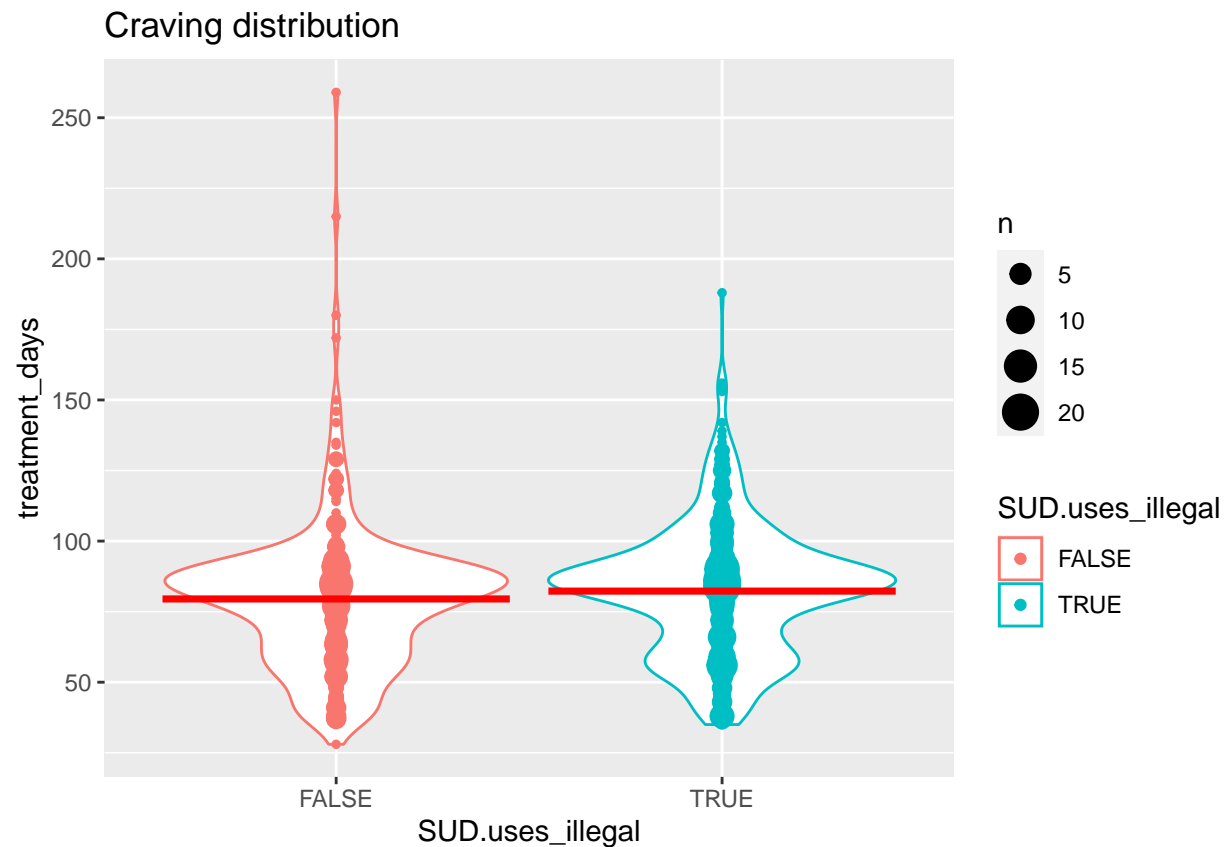
Alcohol/Drug craving baseline to discharge



```
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")
```



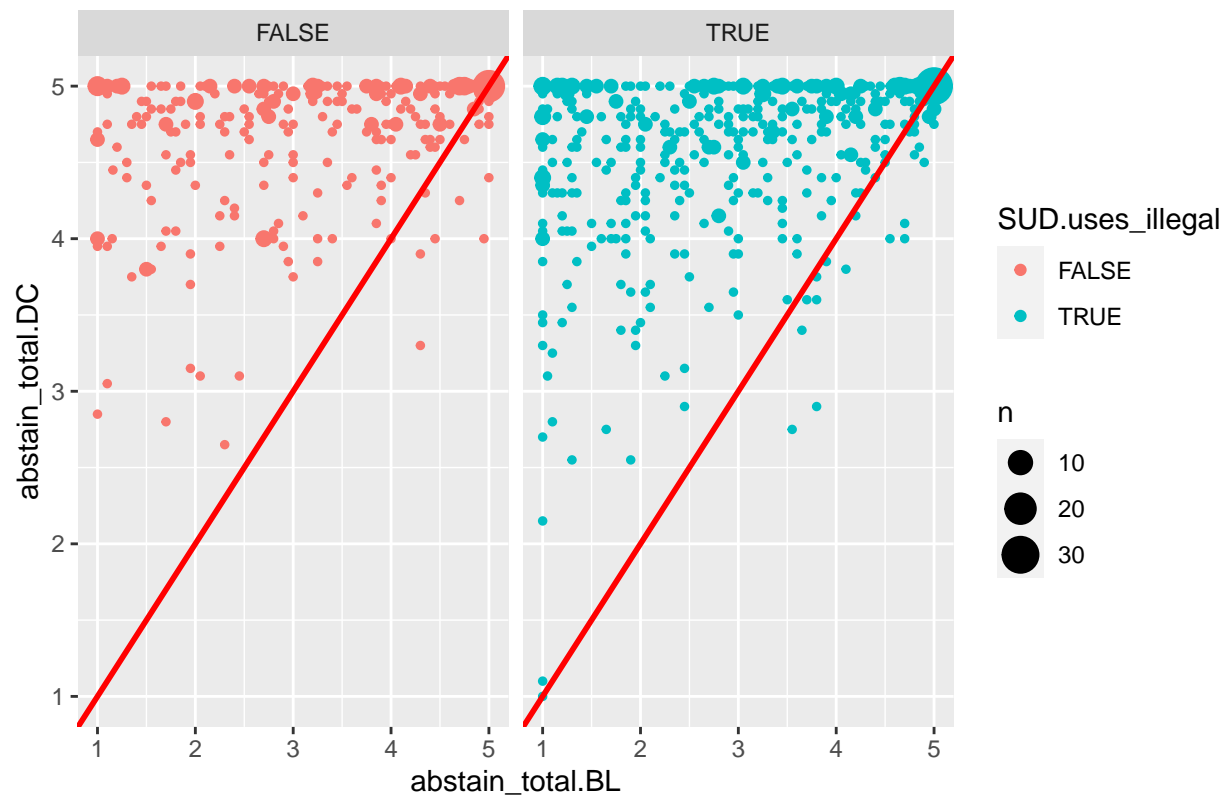
```
cleaned_addiction_data |>
  ggplot(aes(SUD.uses_illegal, treatment_days, color=SUD.uses_illegal)) +
  geom_violin() +
  # geom_jitter(height = 0, width = 0.1) +
  geom_count() +
  stat_summary(fun = "mean",
    geom = "crossbar",
    color = "red") +
  # geom_boxplot(width=0.1) +
  labs(title = "Craving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



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

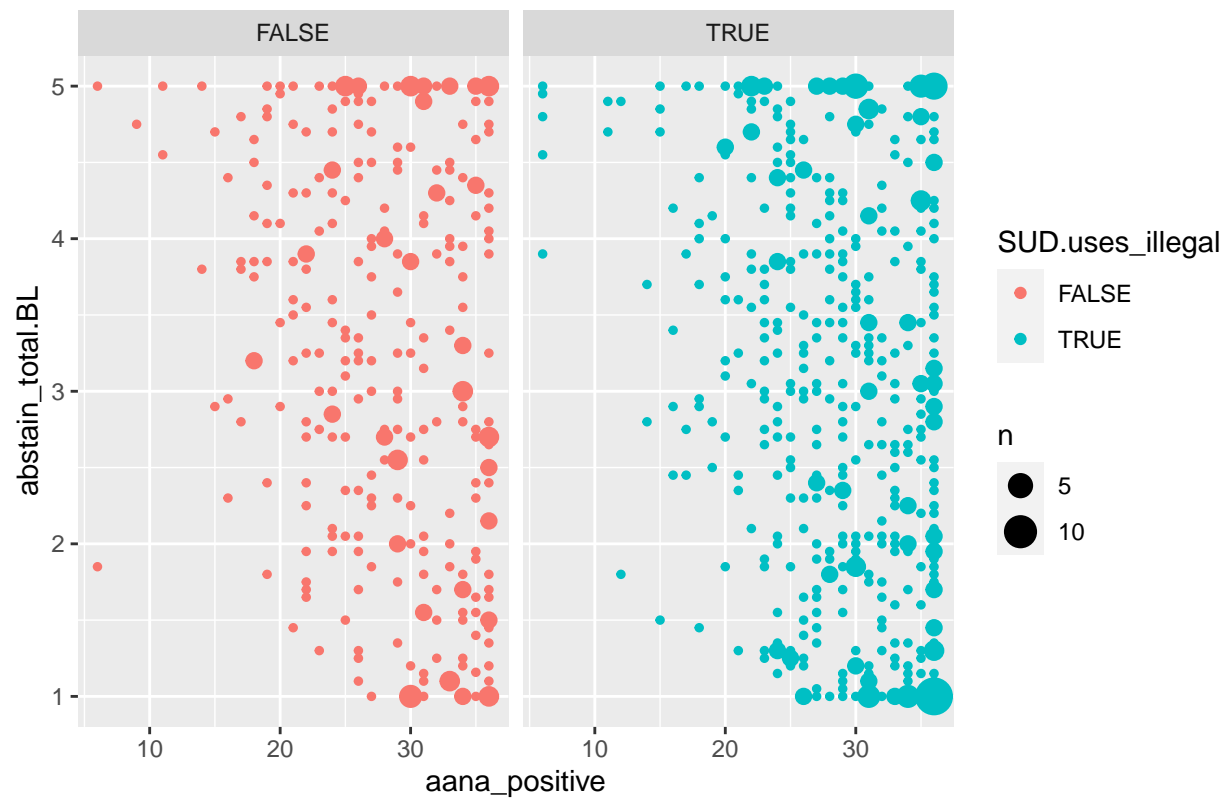
```
cleaned_addiction_data |>
  ggplot(aes(abstain_total.BL, abstain_total.DC, color=SUD.uses_illegal)) +
  geom_count() +
  # scale_x_continuous(n.breaks=20) +
  # scale_y_continuous(n.breaks=20) +
  geom_abline(slope=1, intercept = 0, color="red", size=1, alpha=3) +
  labs(title = "Abstain confidence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


Abstain confidence baseline to discharge



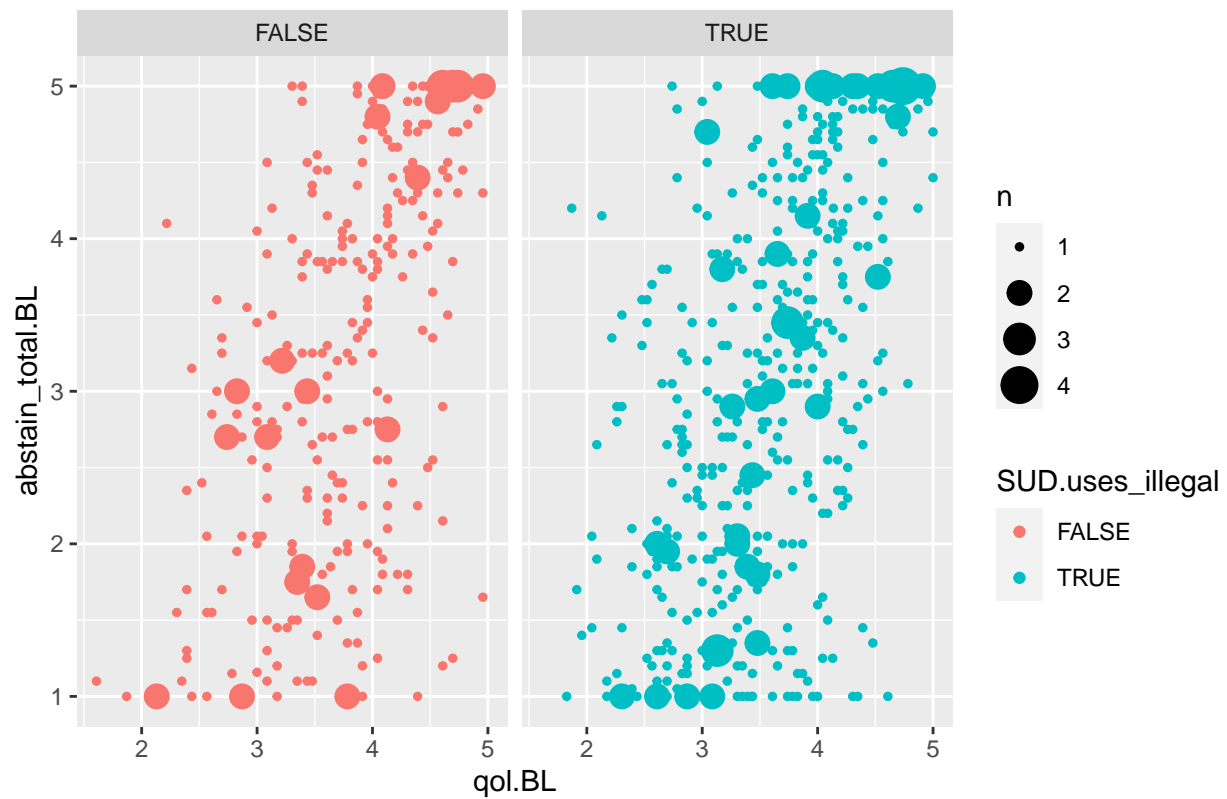
```
cleaned_addiction_data |>
  ggplot(aes(aana_positive, abstain_total.BL, color=SUD.uses_illegal)) +
  geom_count() +
  # scale_x_continuous(n.breaks=20) +
  # scale_y_continuous(n.breaks=20) +
  labs(title = "AA/NA positive regard and abstain confidence") +
  facet_wrap(vars(SUD.uses_illegal))
```

AA/NA positive regard and abstain confidence



```
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))
```

Quality of life and abstain confidence



```
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))
```

Abstain confidence baseline to discharge



```
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

