

# Final Project

Emely Gazarov, Ryo Iwata, Maria Ramirez

24 April, 2024

```
cleaned_addiction_data = read_rds("./addiction.rds") |>
  # mutate(commit_change = (commit.DC - commit.BL) / (commit.BL + commit.DC))
  mutate(commit_change = (commit.DC - commit.BL))

cleaned_addiction_data <- cleaned_addiction_data |>
  mutate(no_social_qol.BL = (qol_general_health.BL + qol_psych.BL + qol_physical.BL + qol_env.BL)/23)

cleaned_addiction_data

## # A tibble: 730 x 173
##   record_id age gender education sober_days treatment_days dropout_yn
##   <dbl> <dbl> <fct> <fct> <dbl> <dbl> <fct>
## 1      87   41 women High School/GED ~ 244 116 no
## 2      94   67 men Bachelor's 17 99 no
## 3      96   28 women Associate's/Some~ 8 92 no
## 4      98   50 men Bachelor's 24 93 no
## 5      99   40 men Associate's/Some~ 6 57 no
## 6     104   35 men Associate's/Some~ 7 59 no
## 7     108   33 women Associate's/Some~ 3 87 no
## 8     109   40 women Associate's/Some~ 7 102 no
## 9     111   59 women High School/GED ~ 7 94 no
## 10    112   39 women High School/GED ~ 5 79 no
## # i 720 more rows
## # i 166 more variables: SUD_is_Alcohol <fct>, SUD_is_Opioid <fct>,
## # SUD_is_Cannabis <fct>, SUD_is_depressants_anxiolytic <fct>,
## # SUD_is_Cocaine <fct>, SUD_is_Other_stimulant <fct>,
## # SUD_is_Hallucinogen <fct>, SUD_is_Nicotine <fct>, SUD_is_Inhalant <fct>,
## # SUD_is_psychosocial <fct>, SUD.sum <dbl>, SUD.sum_legal <dbl>,
## # SUD.uses_legal <fct>, SUD.sum_illegal <dbl>, SUD.uses_illegal <fct>, ...
```

## Overview of what we're working with

### Things that can predict

- number of sober days
- days since baseline
- age
- gender
- education
- drop out yes no
- SUD is alcohol
- SUD is other
- Social support

- family
  - friends
  - sig other
  - total
- Substance use history
  - Tried tobacco/alcohol
  - Age of first use
  - Regular use
  - Age of regular use
- AA/NA affiliation
  - Lifetime number of meetings
  - Last year number of meetings
  - Degree of affiliation
  - Positive thoughts
  - Negative thoughts
- Stressful life
  - happened
  - witnessed
  - learned about
  - exposed
  - total of all things
- Childhood
- Religion
  - Religious affiliation
  - Positive
  - Negative
- Life quality

### Things we can predict

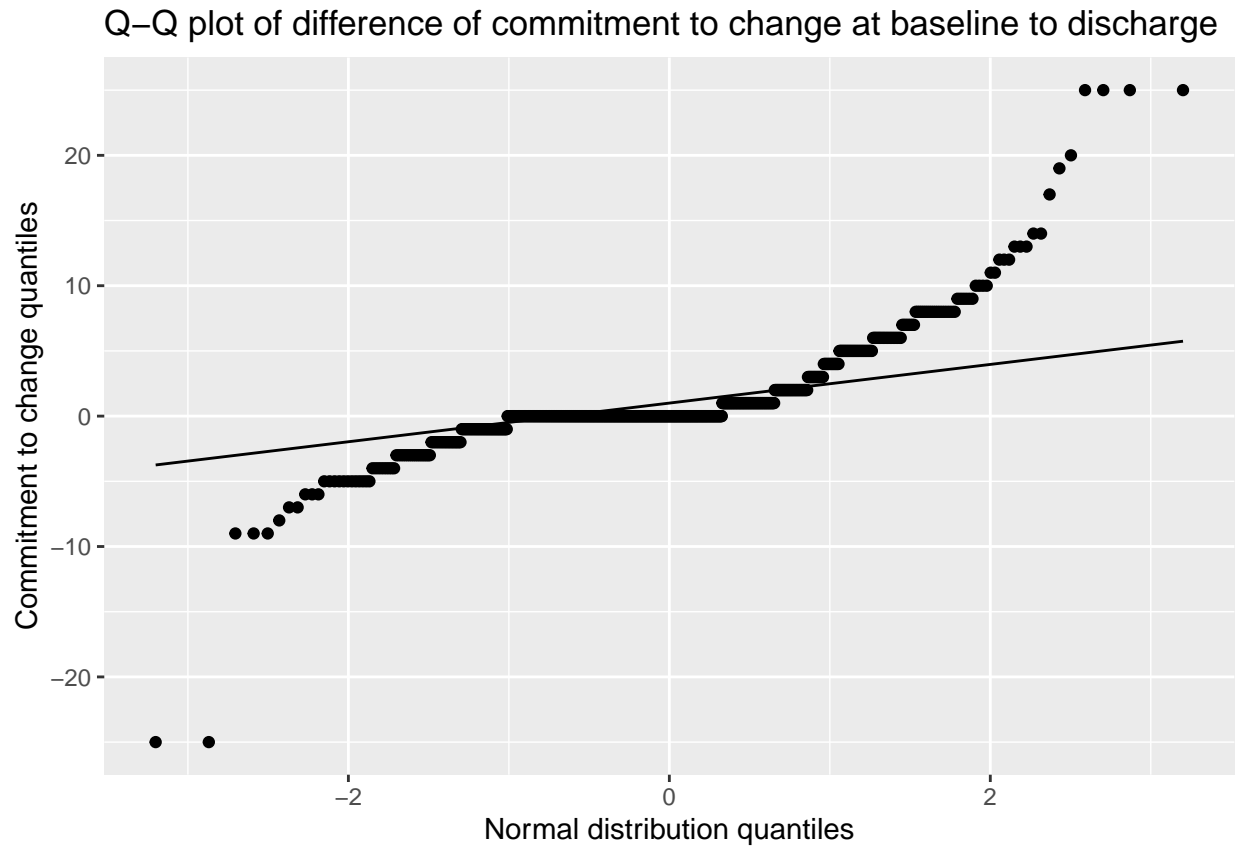
- impression change
- length of stay
- Life quality
- Commitment to change
- Cravings
  - baseline
  - baseline vs followup
- Impression of change

## First hypothesis

```
commitment_model <- lm(commit_change ~ age
  + gender
  + education
  + rel.is_religious
  + religion_pos
  + aana_past_year
  + SUD.sum_illegal
  + SUD.sum_legal
  ,cleaned_addiction_data)
```

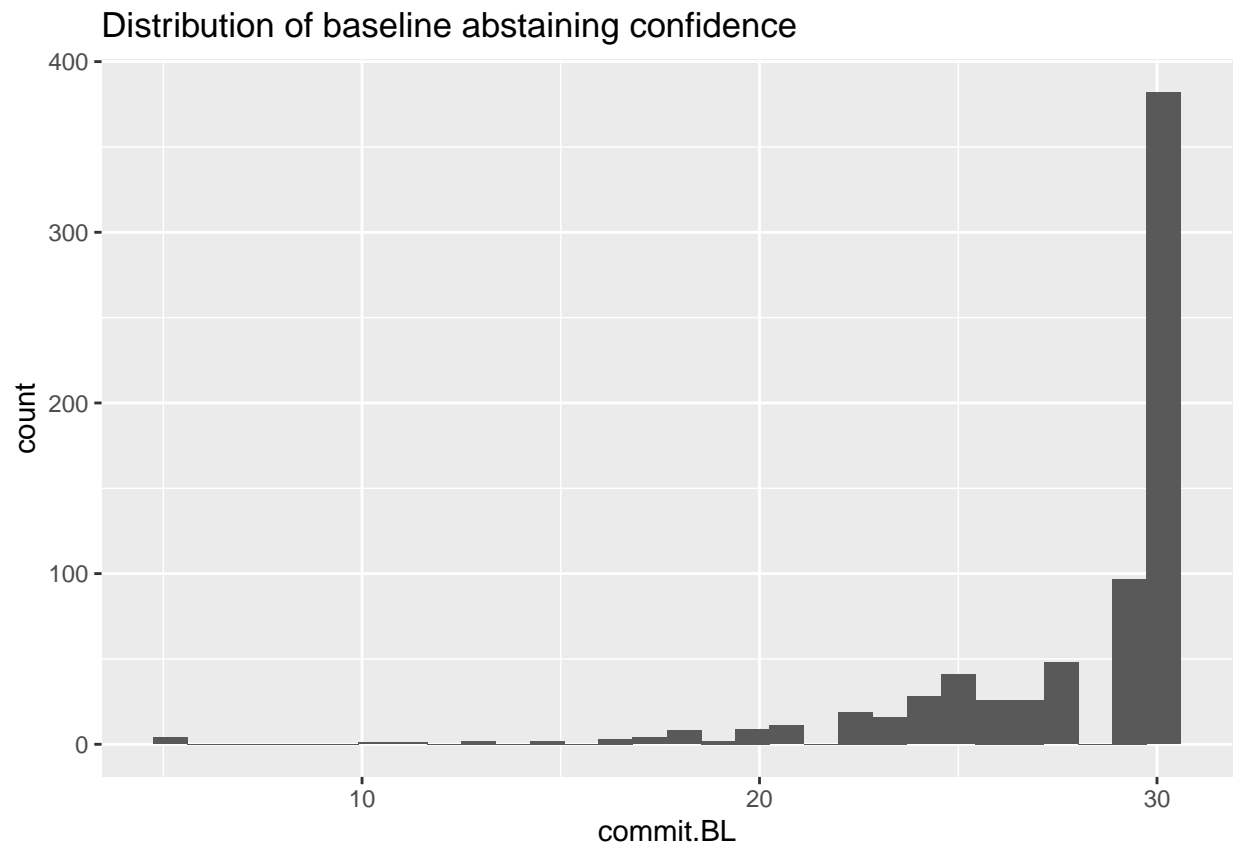
```
# and now the Q-Q plots
cleaned_addiction_data |>
```

```
ggplot(aes(sample=commit_change)) +
  stat_qq() +
  stat_qq_line() +
  labs(
    title = "Q-Q plot of difference of commitment to change at baseline to discharge",
    y = "Commitment to change quantiles",
    x = "Normal distribution quantiles")
```



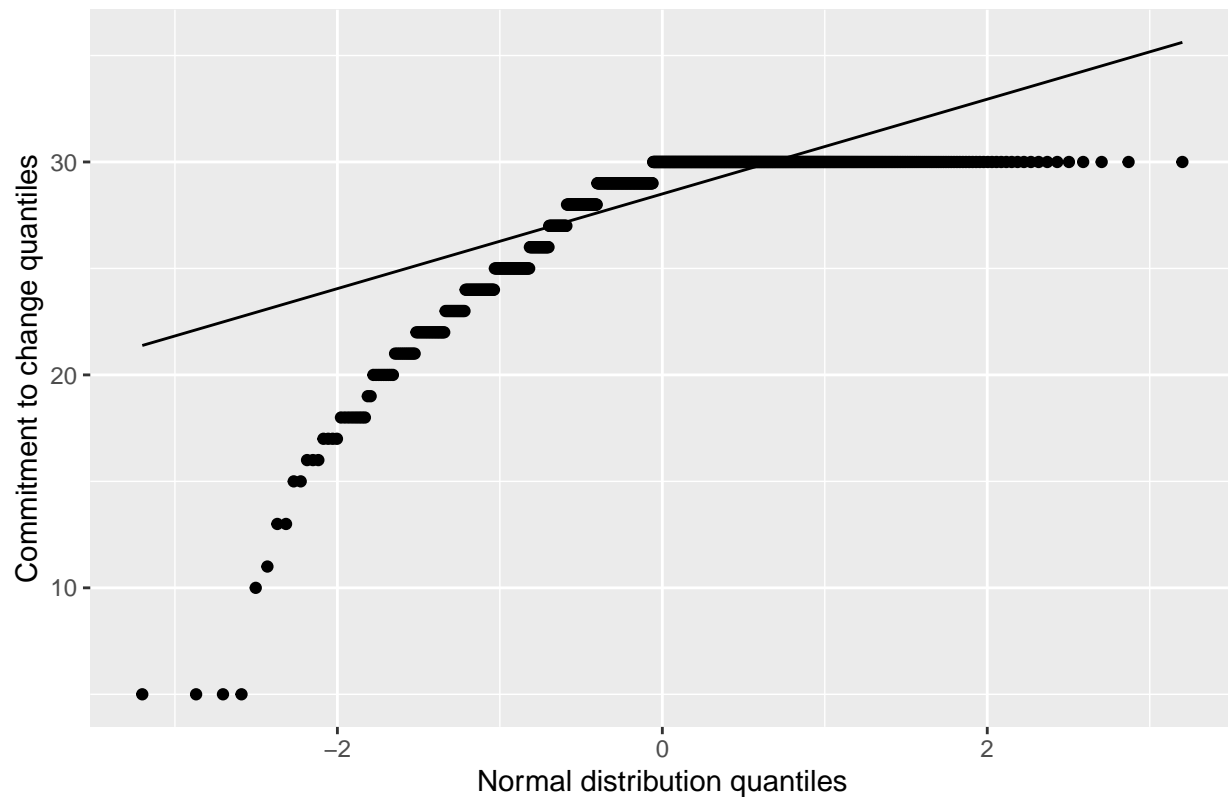
```
# and now the Q-Q plots
cleaned_addiction_data |>
  ggplot(aes(commit.BL)) +
  geom_histogram() +
  labs(
    title = "Distribution of baseline abstaining confidence")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
# and now the Q-Q plots
cleaned_addiction_data |>
  ggplot(aes(sample=commit.BL)) +
  stat_qq() +
  stat_qq_line() +
  labs(
    title = "Q-Q plot of commitment to change at baseline",
    y = "Commitment to change quantiles",
    x = "Normal distribution quantiles")
```

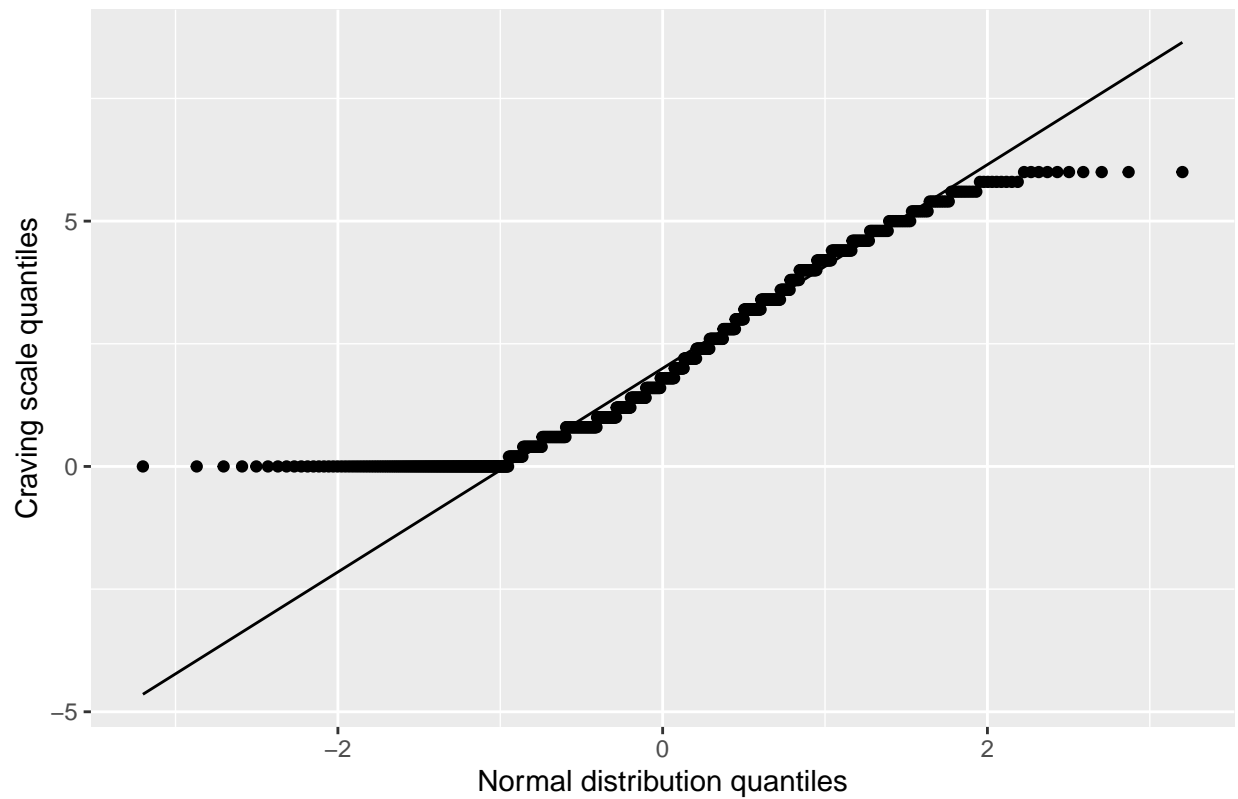
Q-Q plot of commitment to change at baseline



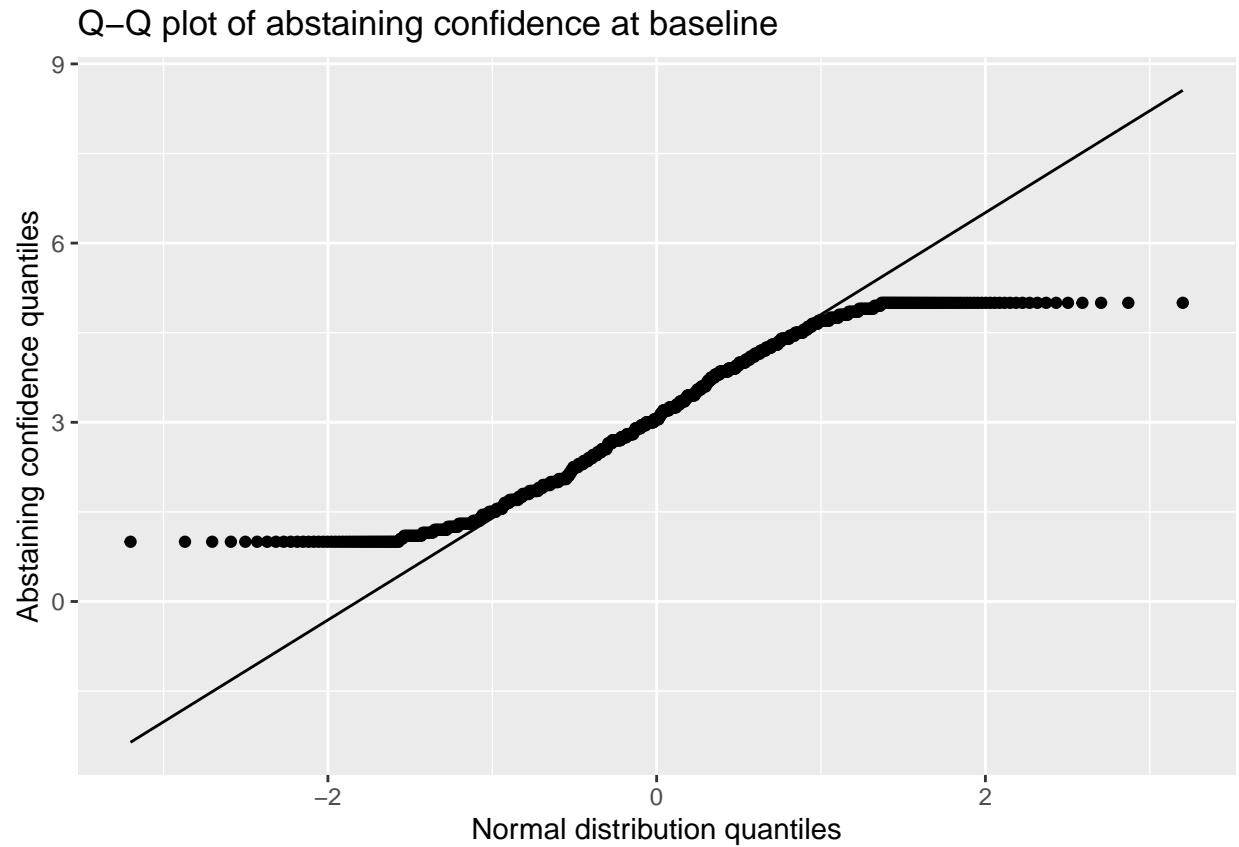
## Looking for other normally distributed things to predict

```
# and now the Q-Q plots
cleaned_addiction_data |>
  ggplot(aes(sample=craving.BL)) +
  stat_qq() +
  stat_qq_line() +
  labs(
    title = "Q-Q plot of craving scale at baseline",
    y = "Craving scale quantiles",
    x = "Normal distribution quantiles")
```

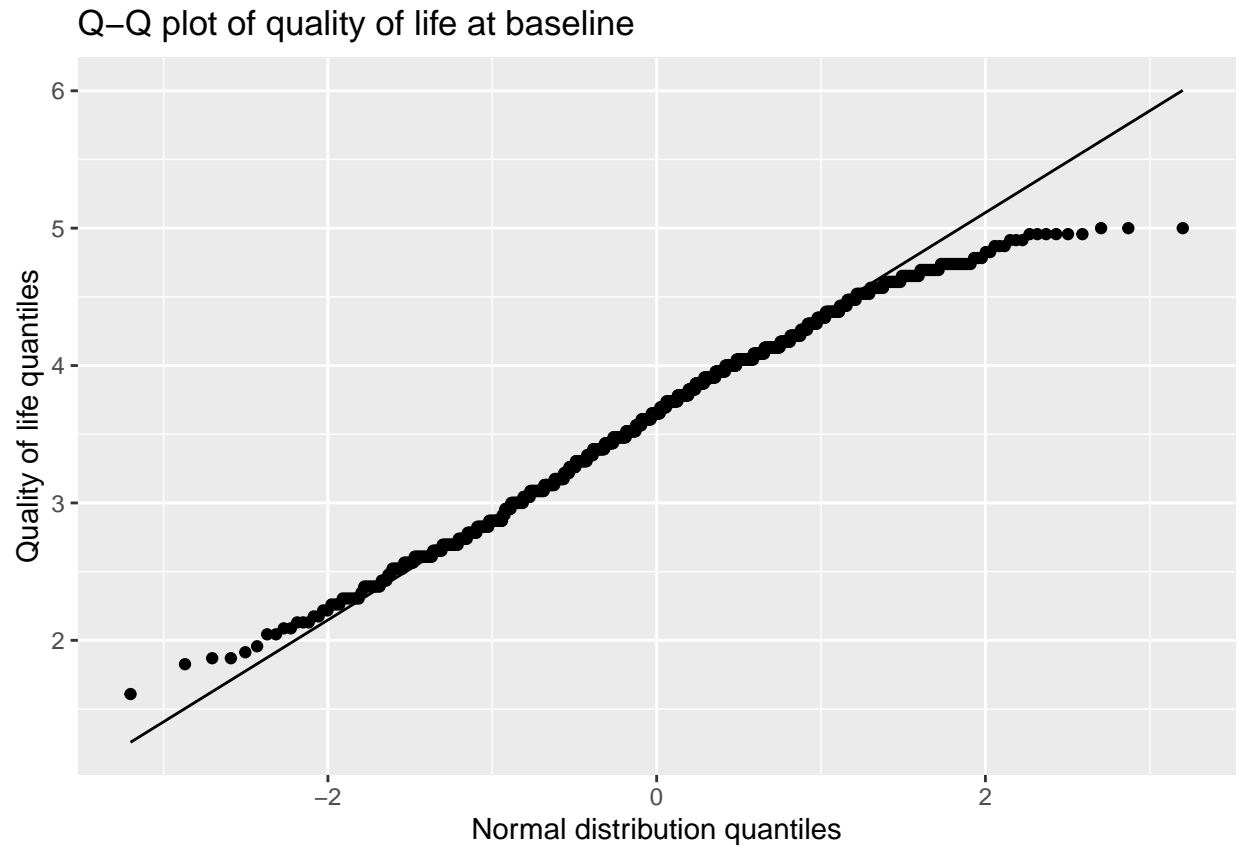
Q-Q plot of craving scale at baseline



```
# and now the Q-Q plots
cleaned_addiction_data |>
  ggplot(aes(sample=abstain_total.BL)) +
  stat_qq() +
  stat_qq_line() +
  labs(
    title = "Q-Q plot of abstaining confidence at baseline",
    y = "Abstaining confidence quantiles",
    x = "Normal distribution quantiles")
```



```
# and now the Q-Q plots
cleaned_addiction_data |>
  ggplot(aes(sample=qol.BL)) +
  stat_qq() +
  stat_qq_line() +
  labs(
    title = "Q-Q plot of quality of life at baseline",
    y = "Quality of life quantiles",
    x = "Normal distribution quantiles")
```

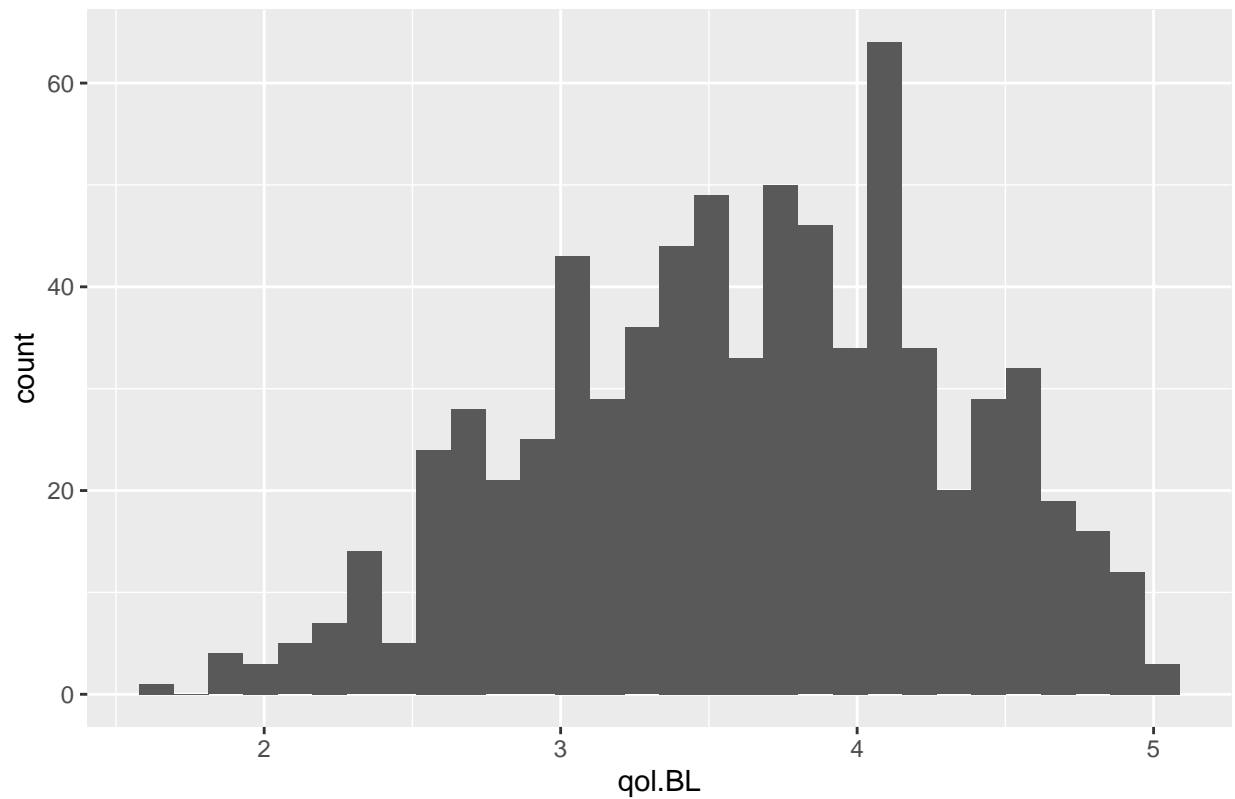


```
# and now the Q-Q plots
cleaned_addiction_data |>
  ggplot(aes(qol.BL)) +
  geom_histogram() +
  labs(
    title = "Distribution of quality of life")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

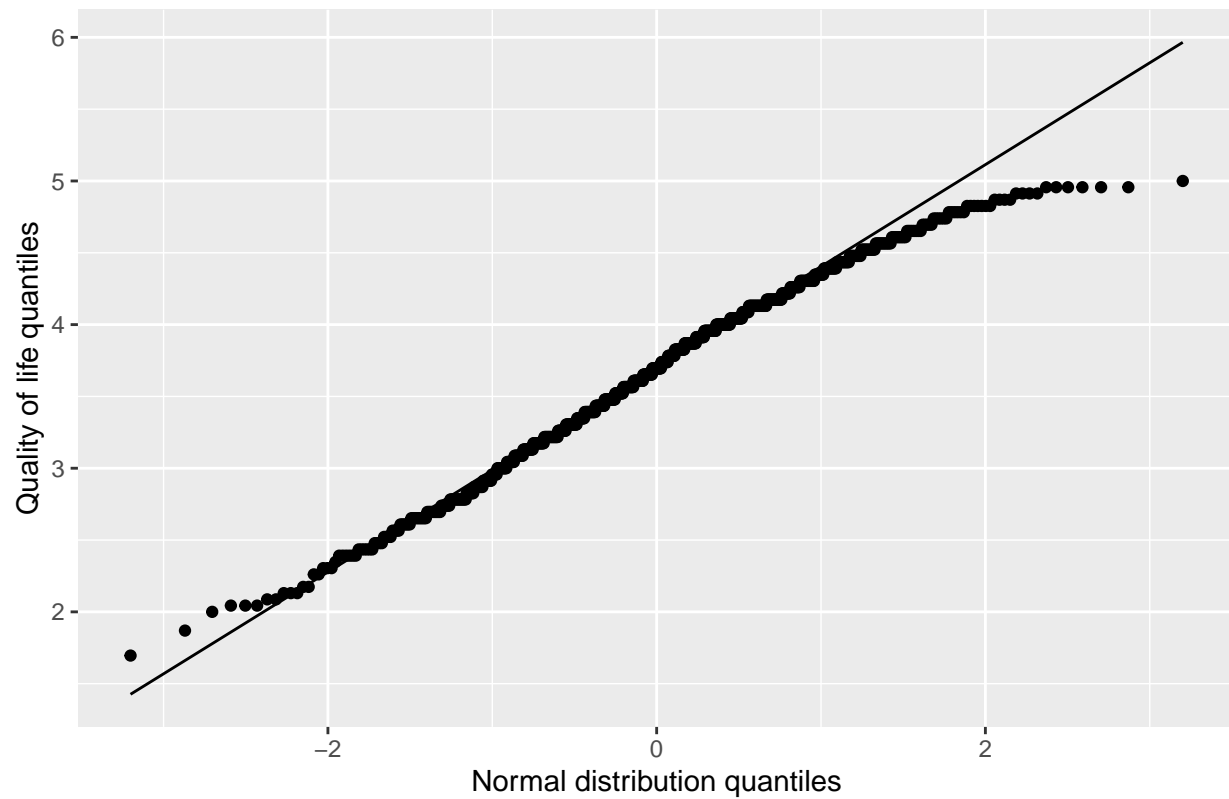


Distribution of quality of life



```
# and now the Q-Q plots
cleaned_addiction_data |>
  ggplot(aes(sample=no_social_qol.BL)) +
  stat_qq() +
  stat_qq_line() +
  labs(
    title = "Q-Q plot of quality of life without social component at baseline",
    y = "Quality of life quantiles",
    x = "Normal distribution quantiles")
```

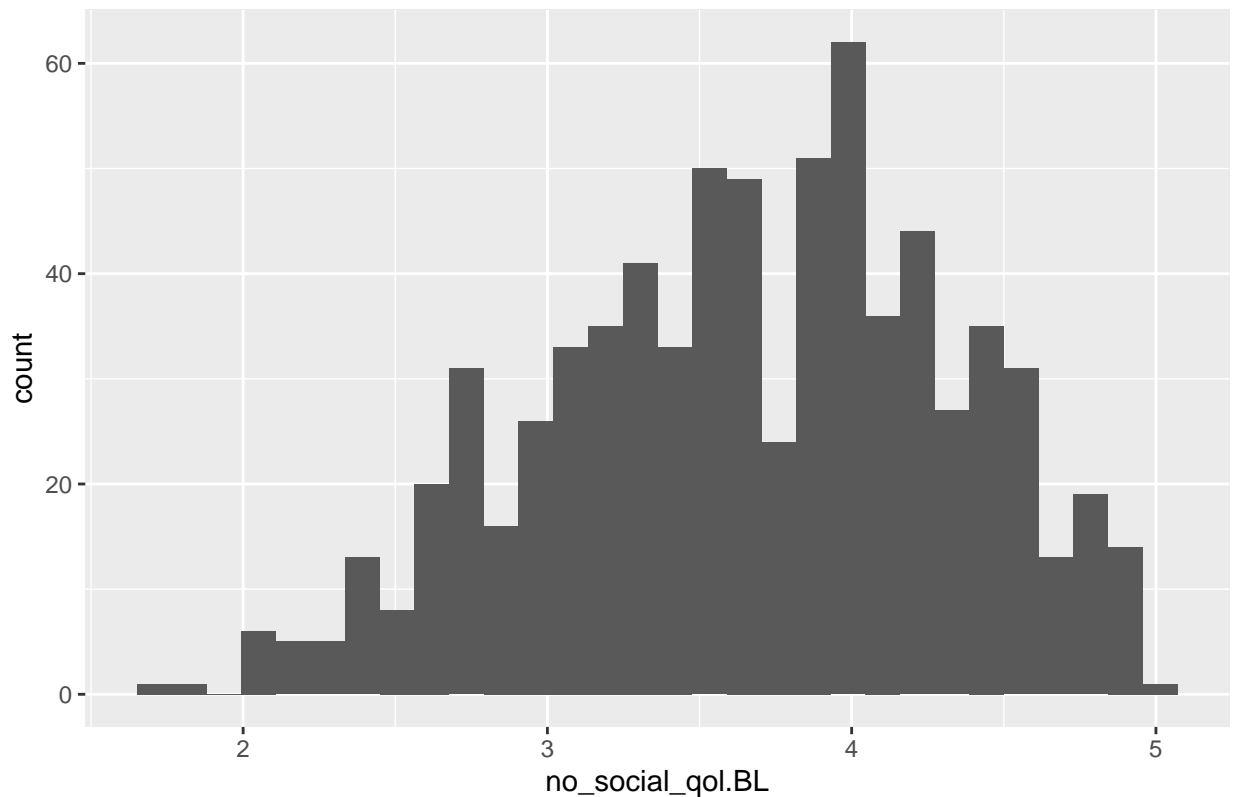
Q-Q plot of quality of life without social component at baseline



```
# and now the Q-Q plots
cleaned_addiction_data |>
  ggplot(aes(no_social_qol.BL)) +
  geom_histogram() +
  labs(
    title = "Distribution of quality of life without social component")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Distribution of quality of life without social component



```
qol_model <- lm(no_social_qol.BL ~ age
  + gender
  + education
  + SUD.sum_illegal
  + SUD.sum_legal
  + rel.is_religious
  + religion_pos
  + aana_past_year
  + aana_positive
  + social
  + stress_to_subj
  ,cleaned_addiction_data)
```

```
summary(qol_model)
```

```
##
## Call:
## lm(formula = no_social_qol.BL ~ age + gender + education + SUD.sum_illegal +
##     SUD.sum_legal + rel.is_religious + religion_pos + aana_past_year +
##     aana_positive + social + stress_to_subj, data = cleaned_addiction_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7143 -0.3570  0.0526  0.3822  1.5135
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.83343    0.18707   15.15 < 2e-16 ***
## age              0.00244    0.00175    1.40  0.1630
## gendermen        0.08085    0.04582    1.76  0.0781 .
## educationAssociate's/Some College -0.07360    0.06284   -1.17  0.2419
## educationBachelor's  0.18656    0.06000    3.11  0.0019 **
## educationMaster's   -0.10187    0.09169   -1.11  0.2669
## educationDoctoral    0.07630    0.07965    0.96  0.3384
## SUD.sum_illegal    -0.04030    0.02017   -2.00  0.0462 *
## SUD.sum_legal      -0.01887    0.02949   -0.64  0.5225
## rel.is_religiousTRUE -0.04346    0.05299   -0.82  0.4124
## religion_pos        0.01671    0.00347    4.81 0.0000018 ***
## aana_past_year0.25  -0.02551    0.04746   -0.54  0.5911
## aana_past_year0.5   -0.01761    0.09531   -0.18  0.8534
## aana_past_year0.75   0.12654    0.11281    1.12  0.2623
## aana_past_year1      0.04958    0.11818    0.42  0.6750
## aana_positive       -0.01472    0.00375   -3.92 0.0000971 ***
## social             0.19089    0.01853   10.30 < 2e-16 ***
## stress_to_subj      -0.03988    0.00891   -4.48 0.0000089 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.561 on 712 degrees of freedom
## Multiple R-squared:  0.304, Adjusted R-squared:  0.287
## F-statistic: 18.3 on 17 and 712 DF, p-value: <2e-16

Anova(qol_model, type=2, digits=4) |>
tidy() |>
kable()

## Warning in printHypothesis(L, rhs, names(b)): one or more coefficients in the hypothesis include
## arithmetic operators in their names;
## the printed representation of the hypothesis will be omitted
```

term	sumsq	df	statistic	p.value
age	0.61383	1	1.95010	0.16301
gender	0.97971	1	3.11249	0.07812
education	7.80550	4	6.19940	0.00007
SUD.sum_illegal	1.25583	1	3.98969	0.04616
SUD.sum_legal	0.12888	1	0.40945	0.52245
rel.is_religious	0.21174	1	0.67269	0.41239
religion_pos	7.29317	1	23.16997	0.00000
aana_past_year	0.67150	4	0.53333	0.71129
aana_positive	4.83695	1	15.36669	0.00010
social	33.41702	1	106.16390	0.00000
stress_to_subj	6.30376	1	20.02669	0.00001
Residuals	224.11497	712	NA	NA

## Looking at model metrics

```
interaction_model <- lm(no_social_qol.BL ~ age
+ gender
+ education
+ SUD.sum_illegal
+ SUD.sum_legal
+ rel.is_religious * religion_pos
+ aana_past_year
+ aana_positive
+ social
+ stress_to_subj
,cleaned_addiction_data)

social_only_model <- lm(no_social_qol.BL ~ age
+ gender
+ education
+ SUD.sum_illegal
+ SUD.sum_legal
+ social
,cleaned_addiction_data)

with_religion_model <- lm(no_social_qol.BL ~ age
+ gender
+ education
+ SUD.sum_illegal
+ SUD.sum_legal
+ social
+ rel.is_religious * religion_pos
,cleaned_addiction_data)

with_aana_model <- lm(no_social_qol.BL ~ age
+ gender
+ education
+ SUD.sum_illegal
+ SUD.sum_legal
+ social
+ aana_past_year
+ aana_positive
,cleaned_addiction_data)

with_stress_model <- lm(no_social_qol.BL ~ age
+ gender
+ education
+ SUD.sum_illegal
+ SUD.sum_legal
+ social
+ stress_to_subj
,cleaned_addiction_data)
```

```
kable(AIC(qol_model, interaction_model, social_only_model, with_aana_model, with_religion_model, with_s
```

	df	AIC
qol_model	19	1247.6
interaction_model	20	1249.5
social_only_model	11	1288.8
with_aana_model	16	1283.6
with_religion_model	14	1279.2
with_stress_model	12	1272.2