

ECDF

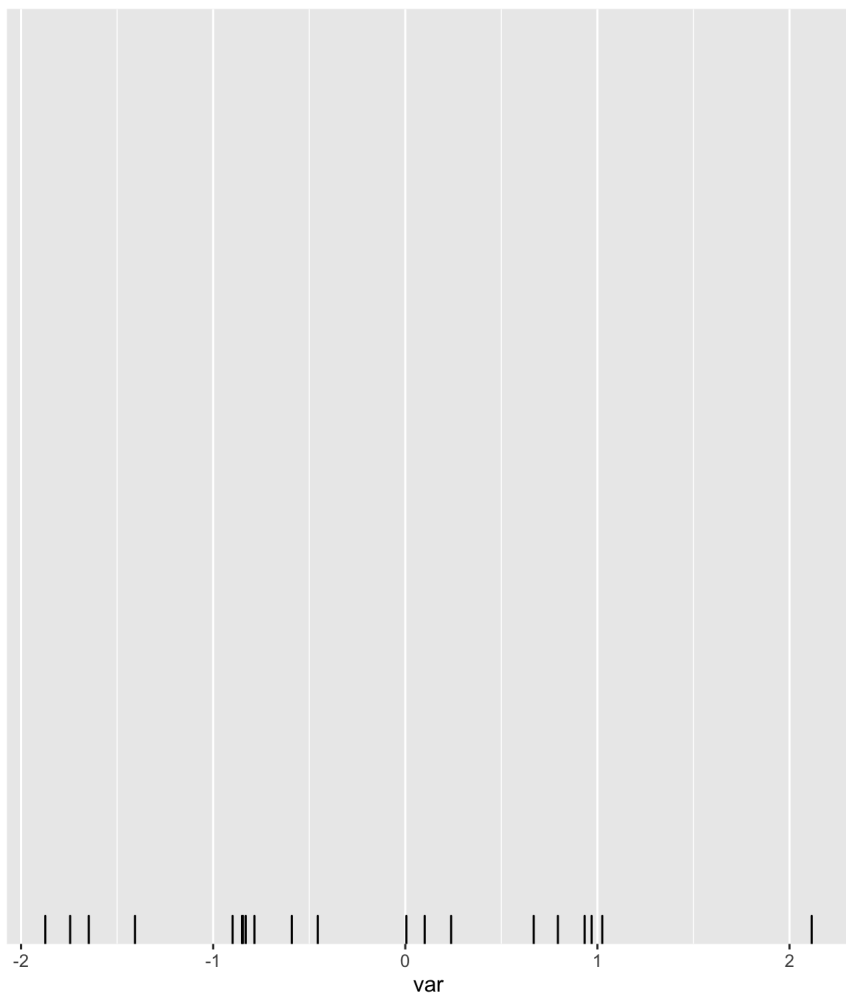
With flipbookr and xaringan

Gina Reynolds

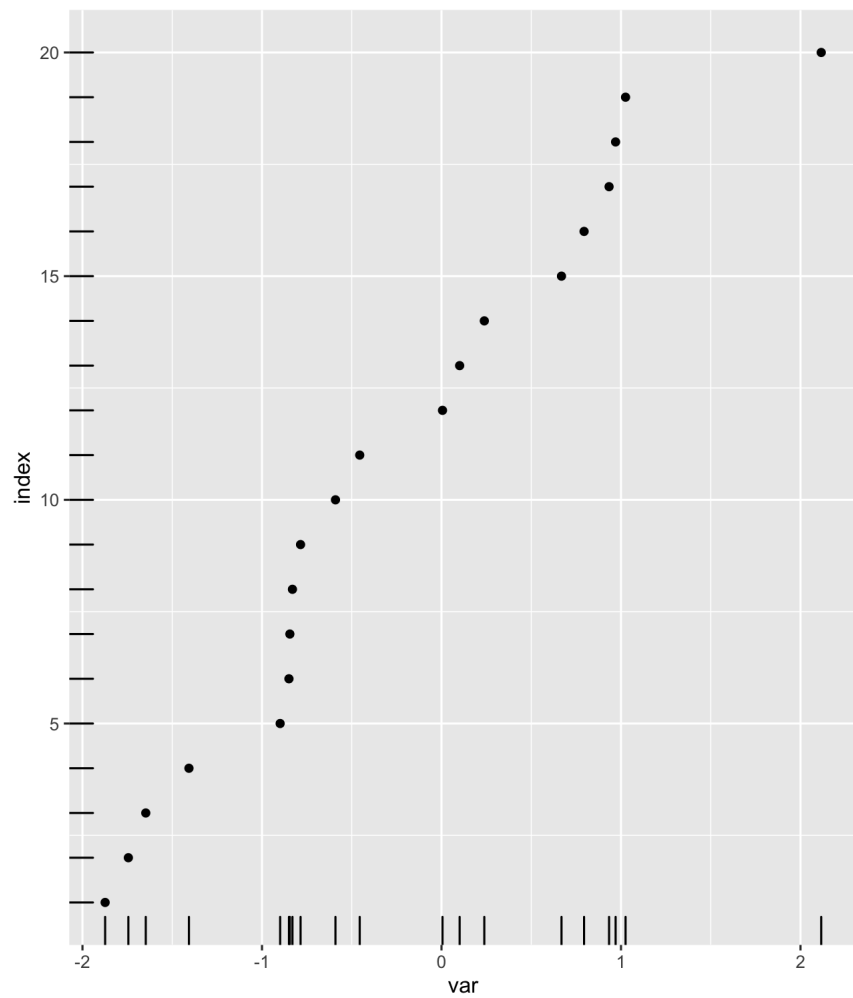

```
library(tidyverse)
```

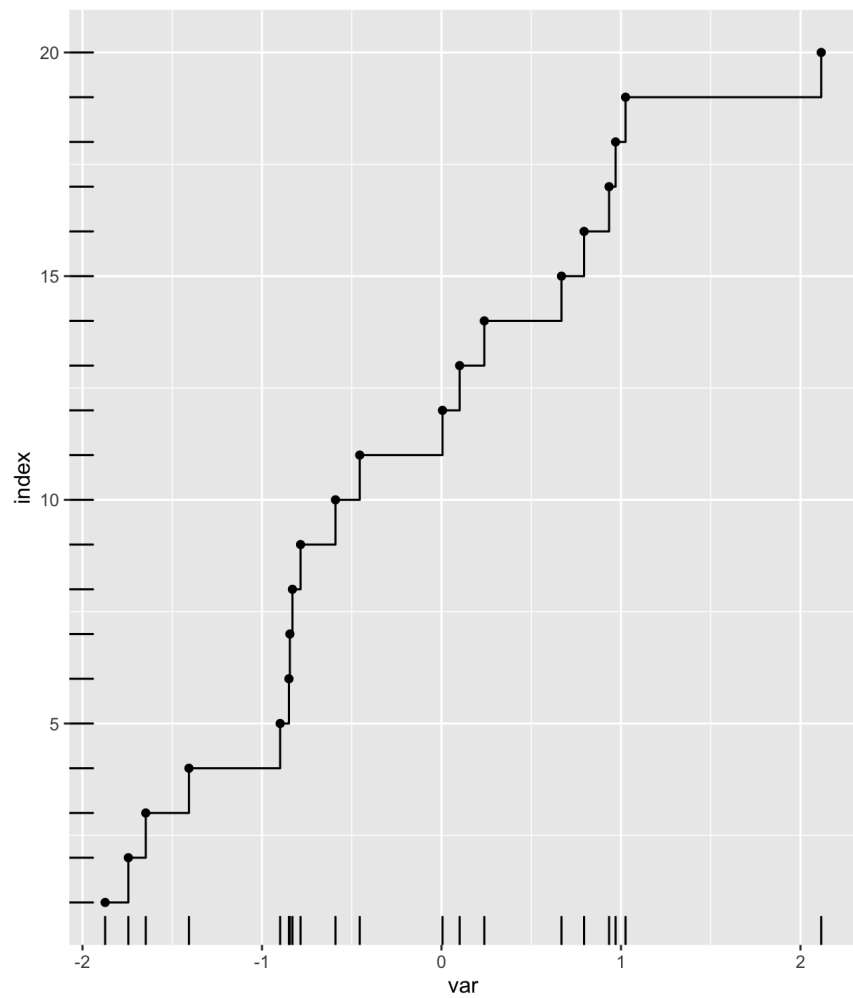
```
library(patchwork)
```

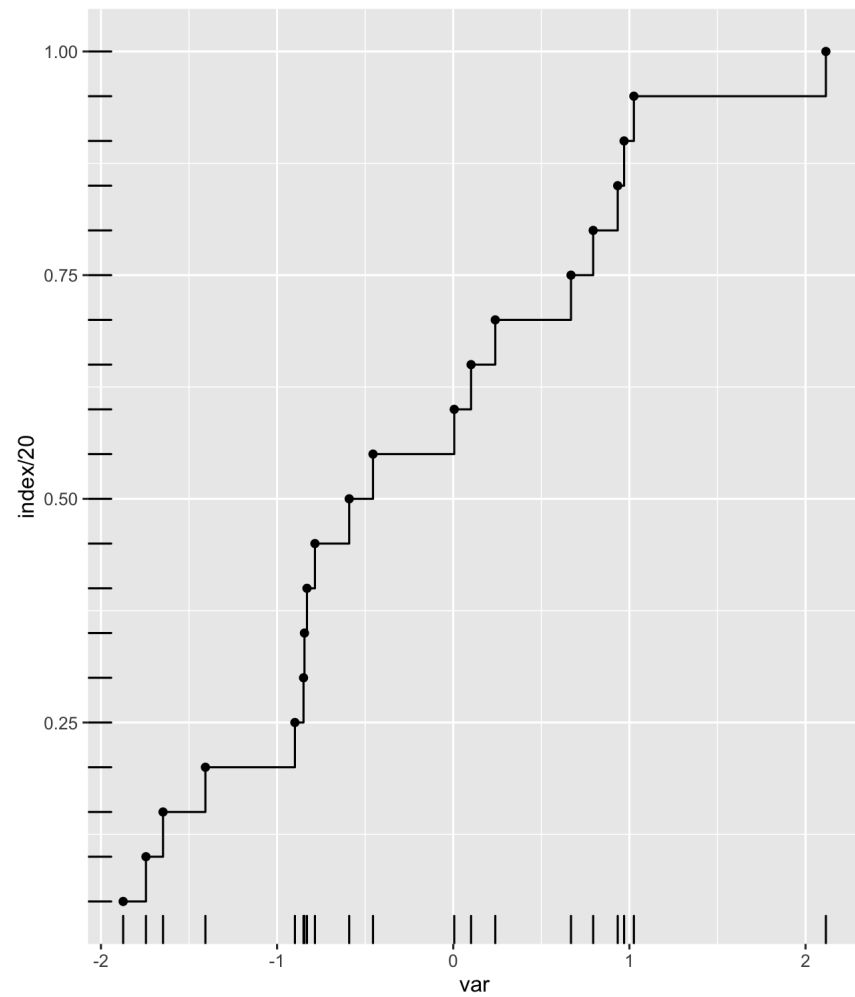
Small N vis walk through

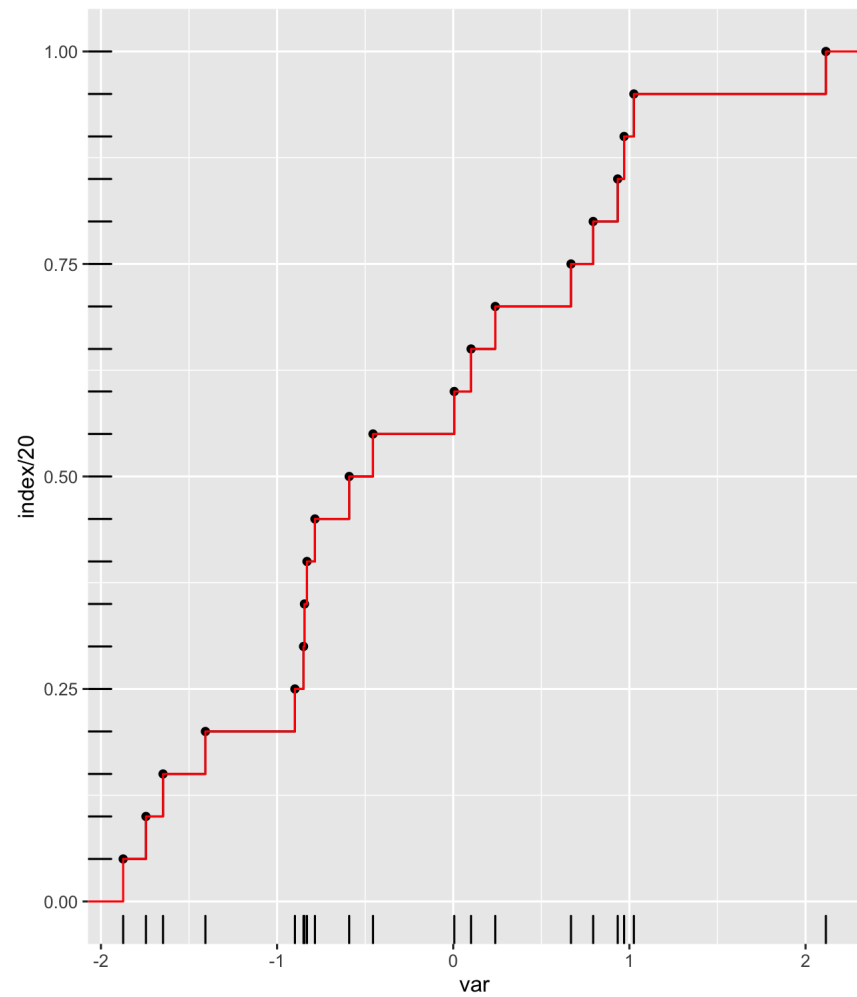


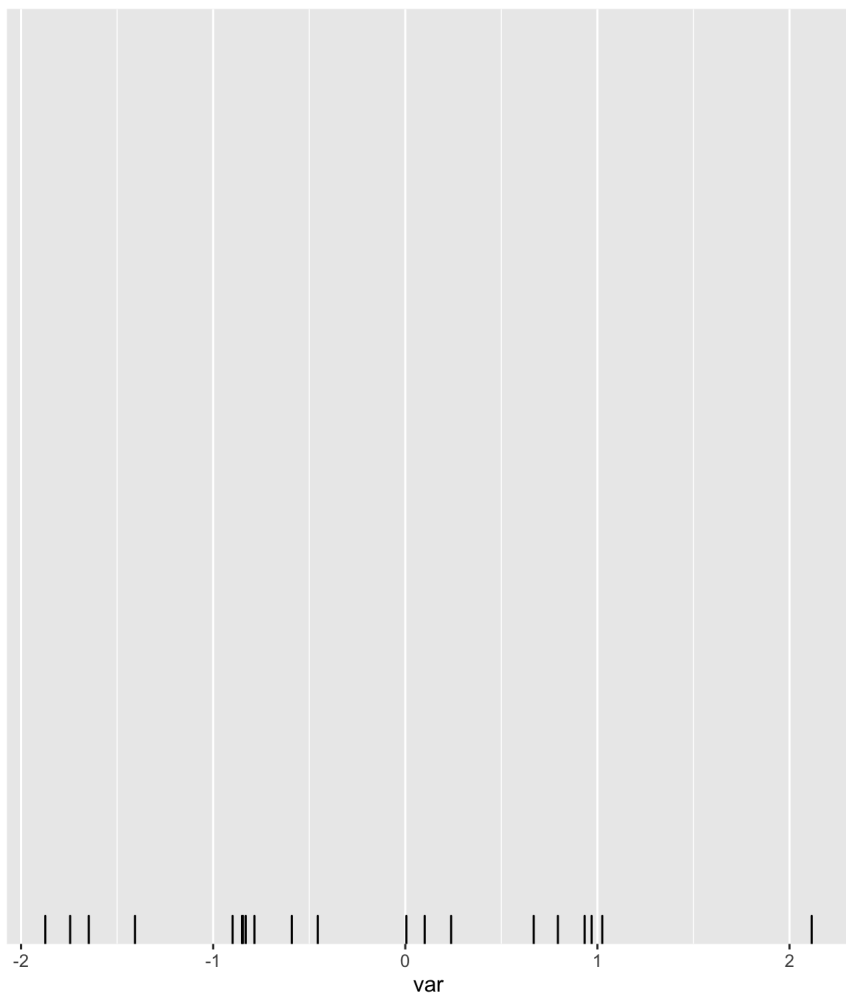


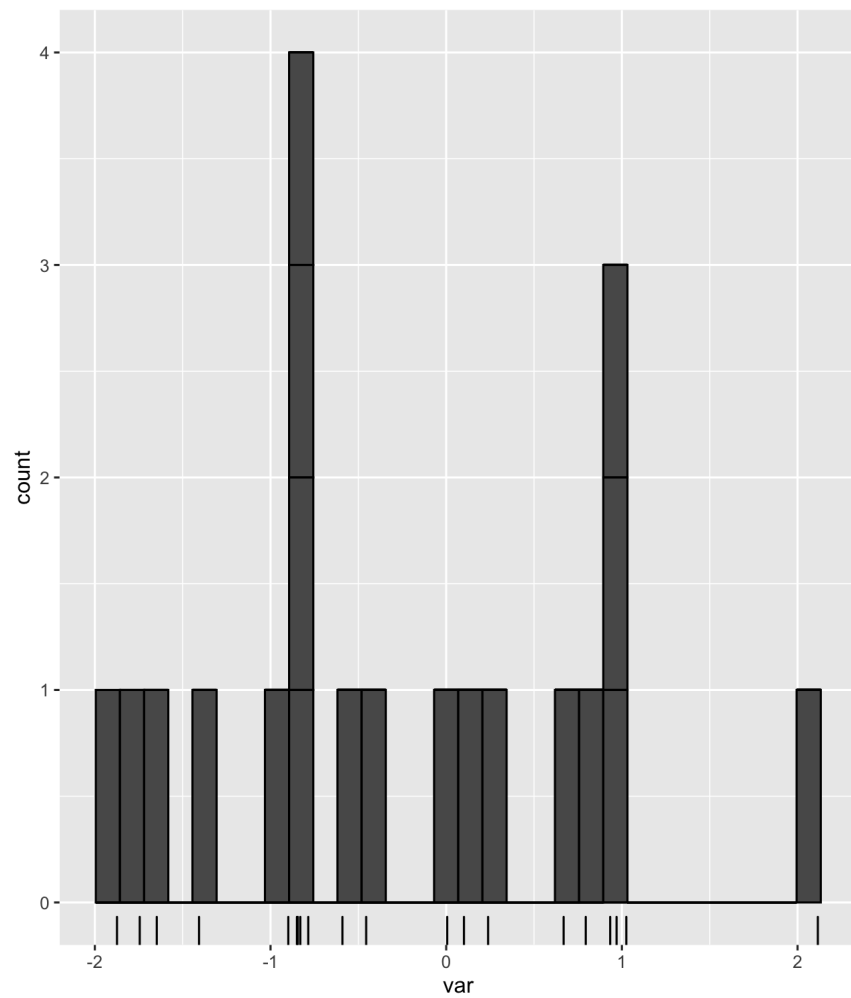


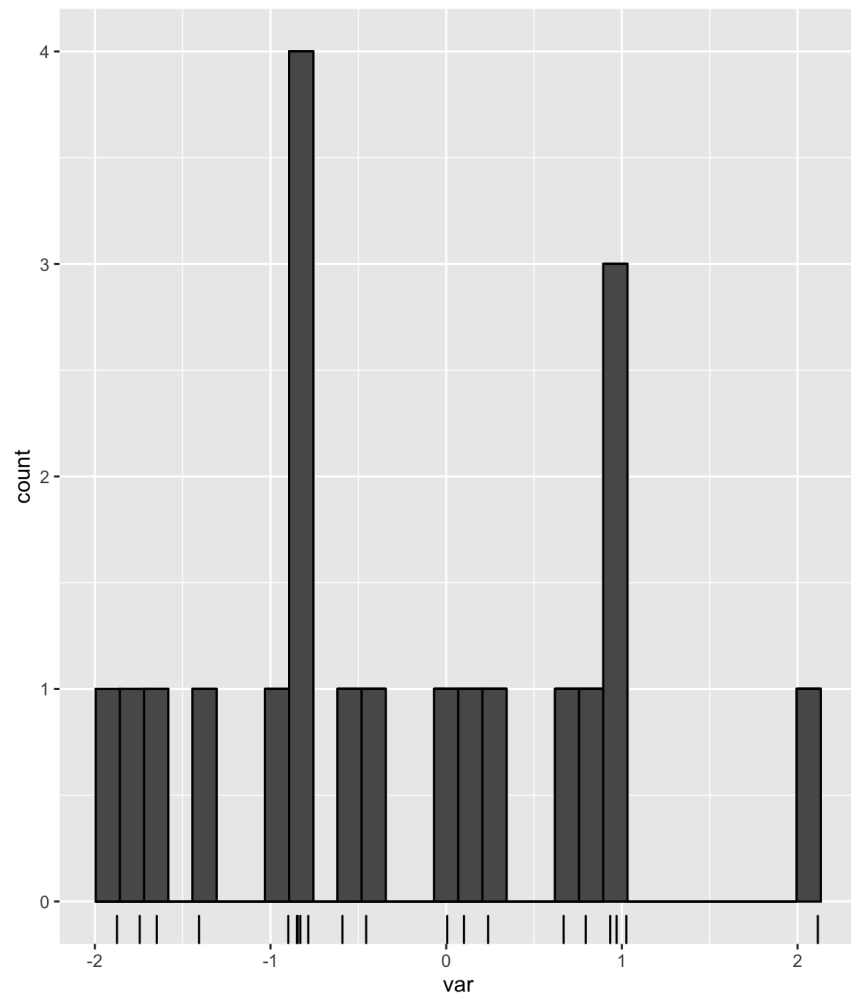


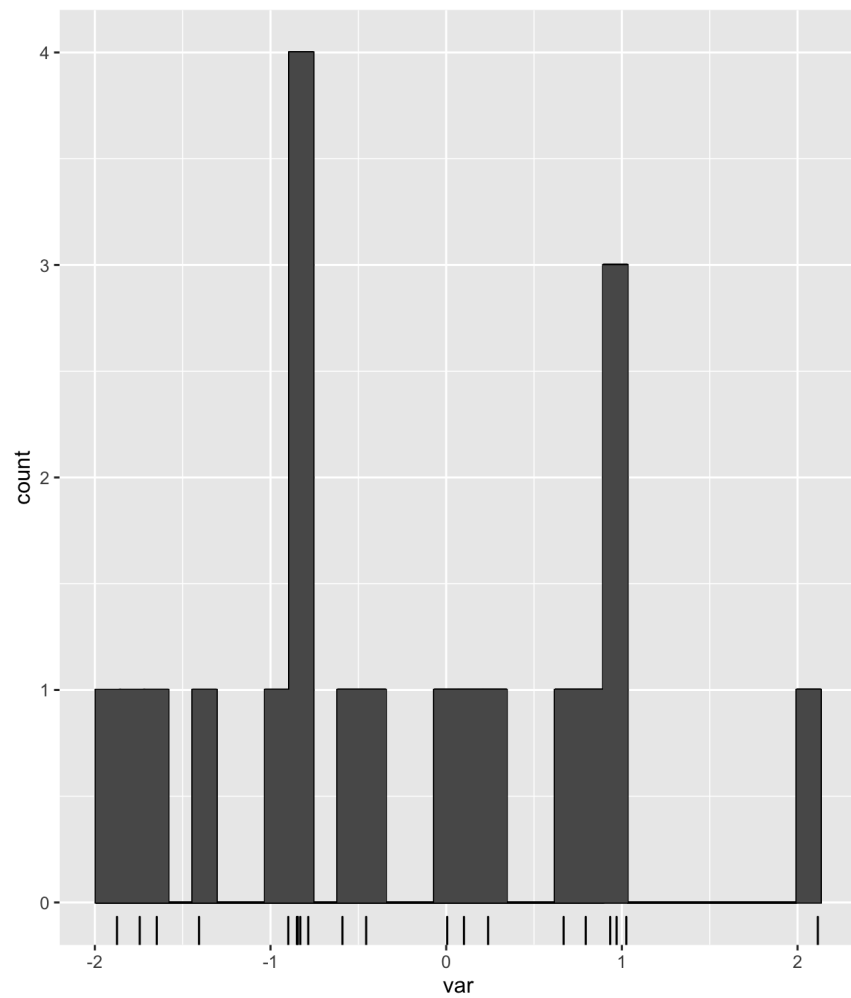


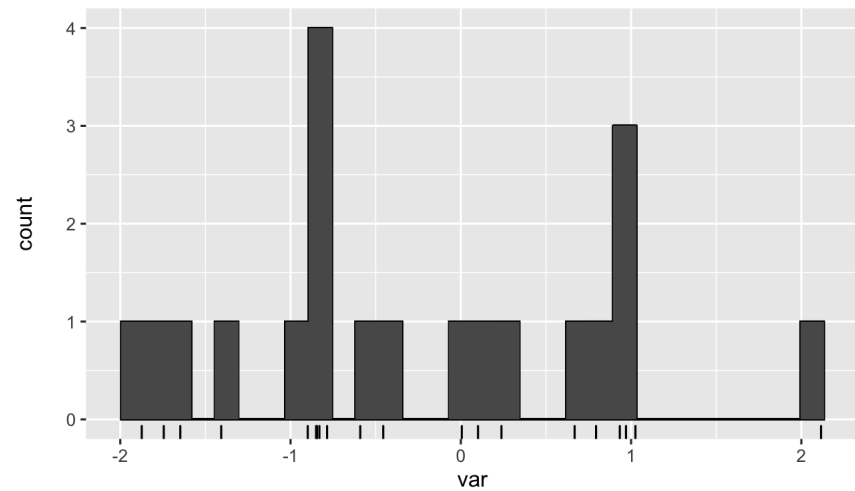
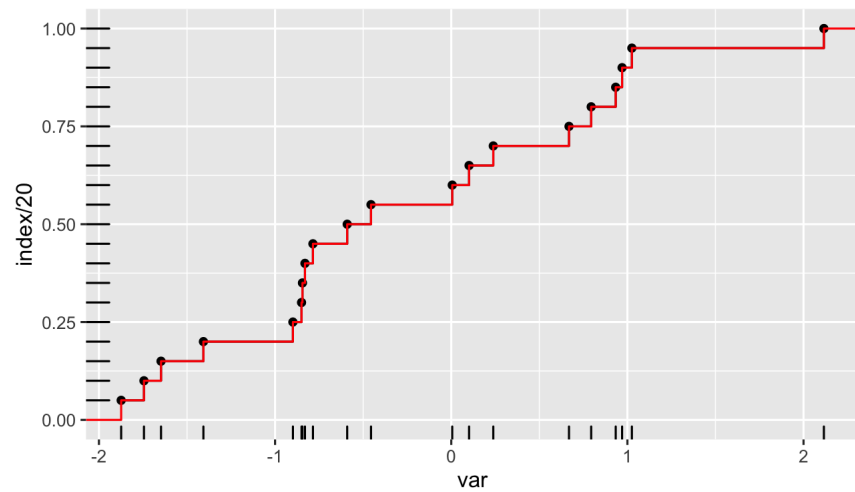












Small N with code/vis walk through


```
set.seed(780)
```

```
set.seed(780)
```

```
rnorm(20)
```

```
[1] -1.647531241 -0.898891586  0.005635507 -0.455534937  2.115503581  
[6]  0.101123587 -1.873883654  0.969946060 -0.830382178  0.933596933  
[11] -0.590622722 -1.406949386  0.238567023 -1.744497088 -0.849863390  
[16]  1.025406106  0.794348885  0.668462054 -0.844740960 -0.785183840
```

```
set.seed(780)
```

```
rnorm(20) %>%
```

```
tibble(var = .)
```

```
# A tibble: 20 x 1
```

```
var
```

```
<dbl>
```

```
1 -1.65
```

```
2 -0.899
```

```
3 0.00564
```

```
4 -0.456
```

```
5 2.12
```

```
6 0.101
```

```
7 -1.87
```

```
8 0.970
```

```
9 -0.830
```

```
10 0.934
```

```
11 -0.591
```

```
12 -1.41
```

```
13 0.239
```

```
14 -1.74
```

```
15 -0.850
```

```
16 1.03
```

```
17 0.794
```

```
18 0.668
```

```
19 -0.845
```

```
20 -0.785
```

```
set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var)
```

```
# A tibble: 20 x 1
      var
  <dbl>
1 -1.87
2 -1.74
3 -1.65
4 -1.41
5 -0.899
6 -0.850
7 -0.845
8 -0.830
9 -0.785
10 -0.591
11 -0.456
12  0.00564
13  0.101
14  0.239
15  0.668
16  0.794
17  0.934
18  0.970
19  1.03
20  2.12
```

```
set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n())
```

```
# A tibble: 20 x 2
      var index
  <dbl> <int>
1 -1.87     1
2 -1.74     2
3 -1.65     3
4 -1.41     4
5 -0.899    5
6 -0.850    6
7 -0.845    7
8 -0.830    8
9 -0.785    9
10 -0.591   10
11 -0.456   11
12  0.00564 12
13  0.101   13
14  0.239   14
15  0.668   15
16  0.794   16
17  0.934   17
18  0.970   18
19  1.03    19
20  2.12    20
```

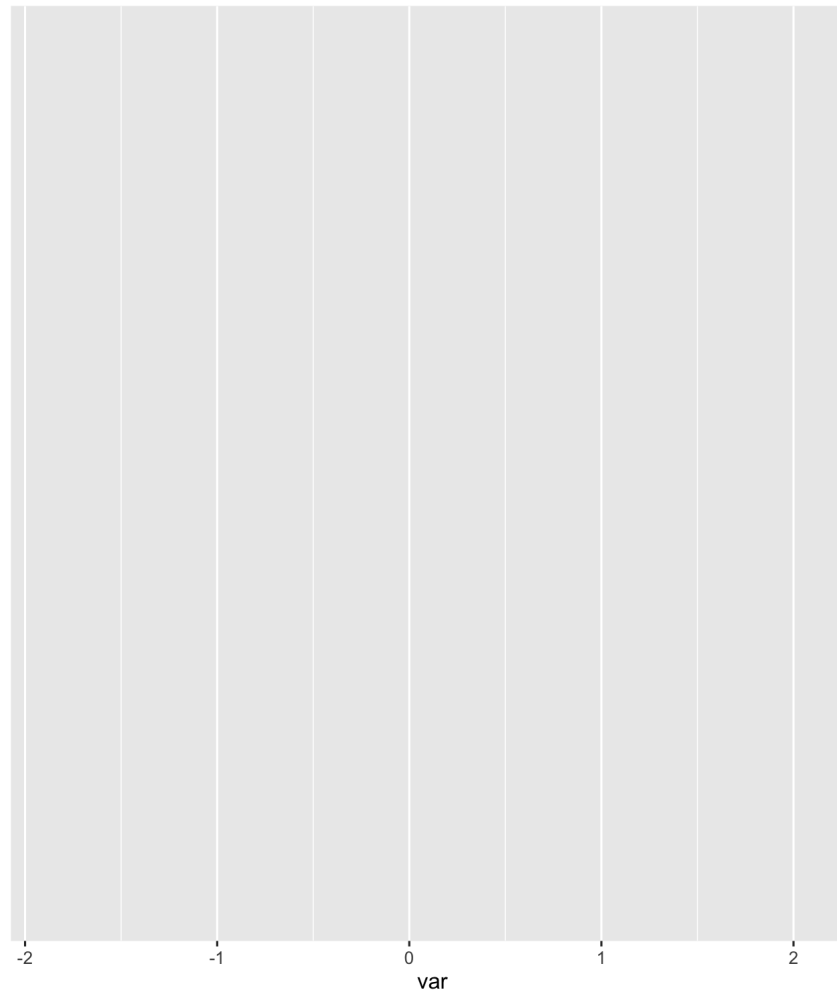
```
set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df
```

```
set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df)
```

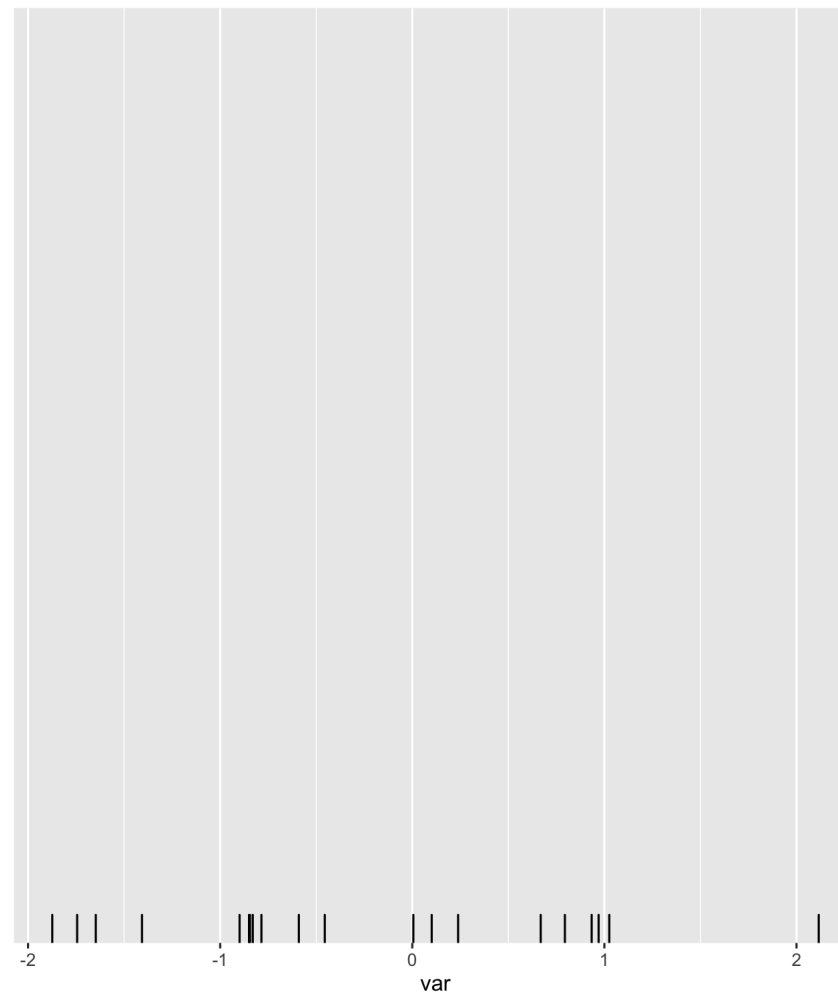
```
set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var)
```



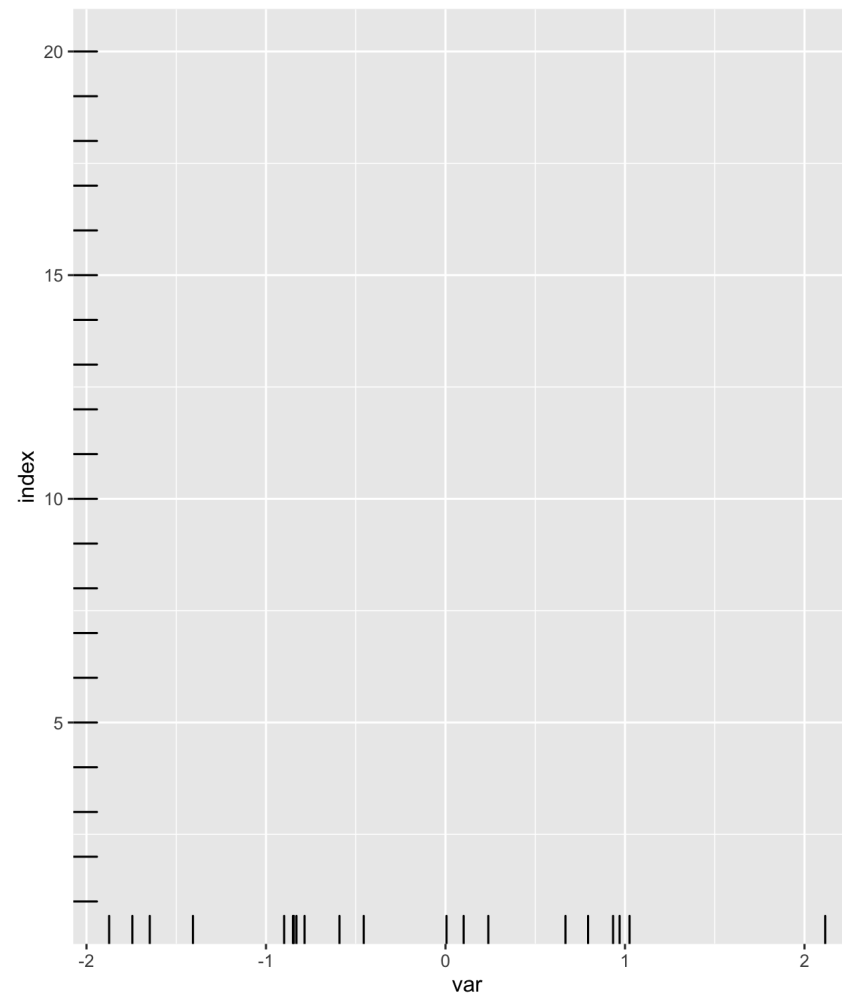

```
set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug()
```



```
set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index)
```

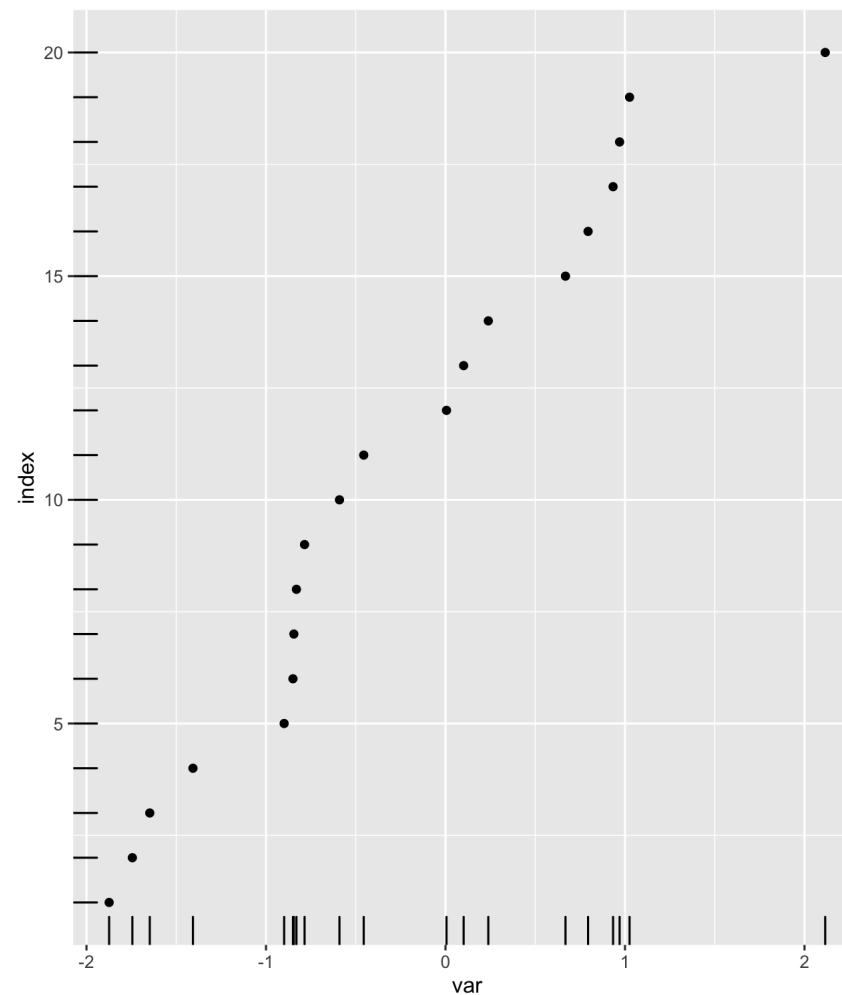


```

set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point()

```

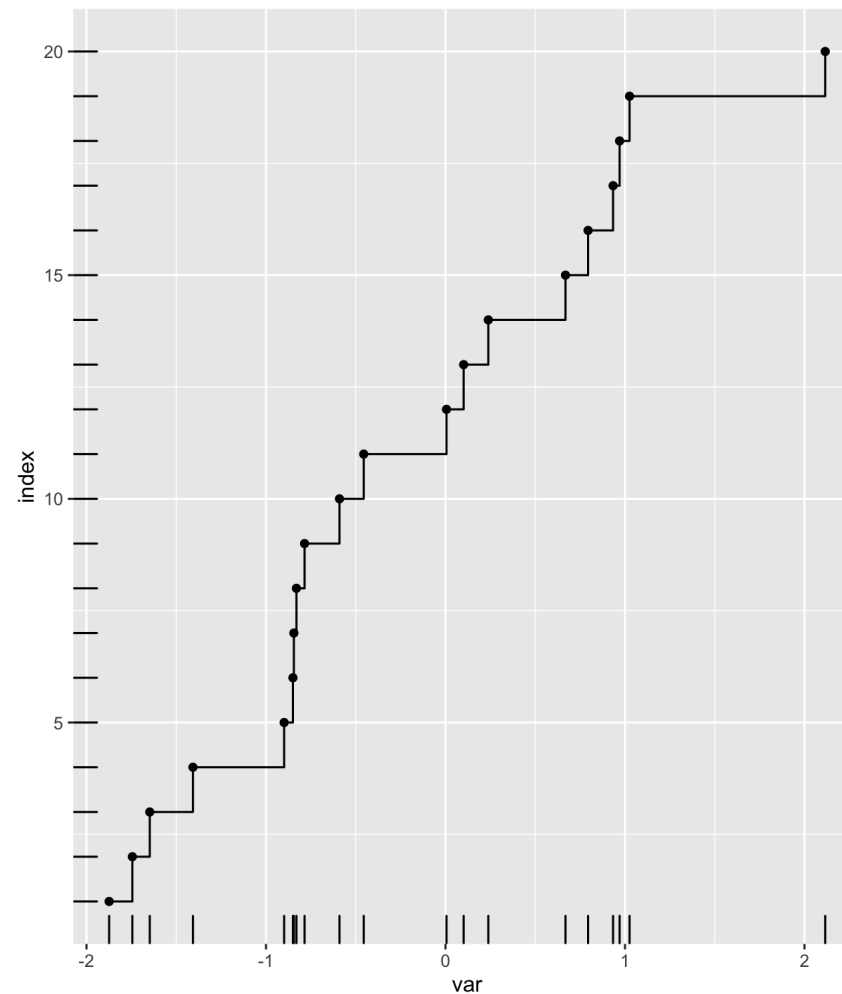


```

set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step()

```

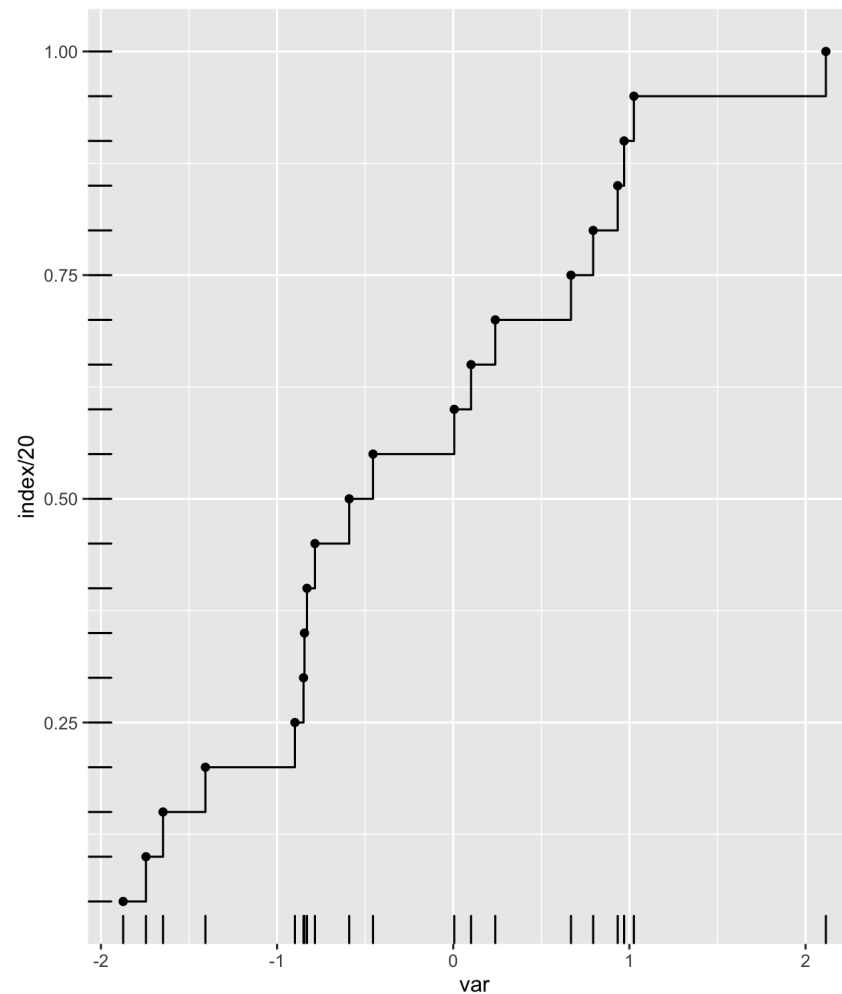


```

set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 20)

```

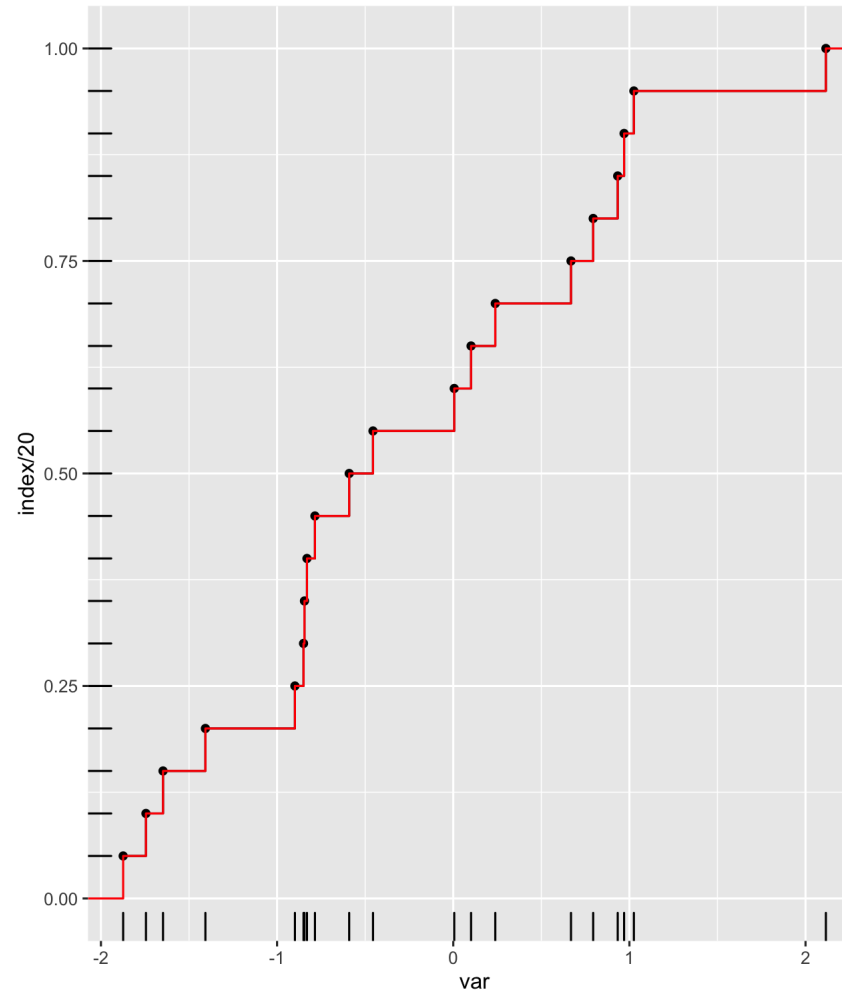


```

set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 20) +
  stat_ecdf(color = "red")

```



```
set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 20) +
  stat_ecdf(color = "red") ->
ecdf
```

```
set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 20) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df)
```



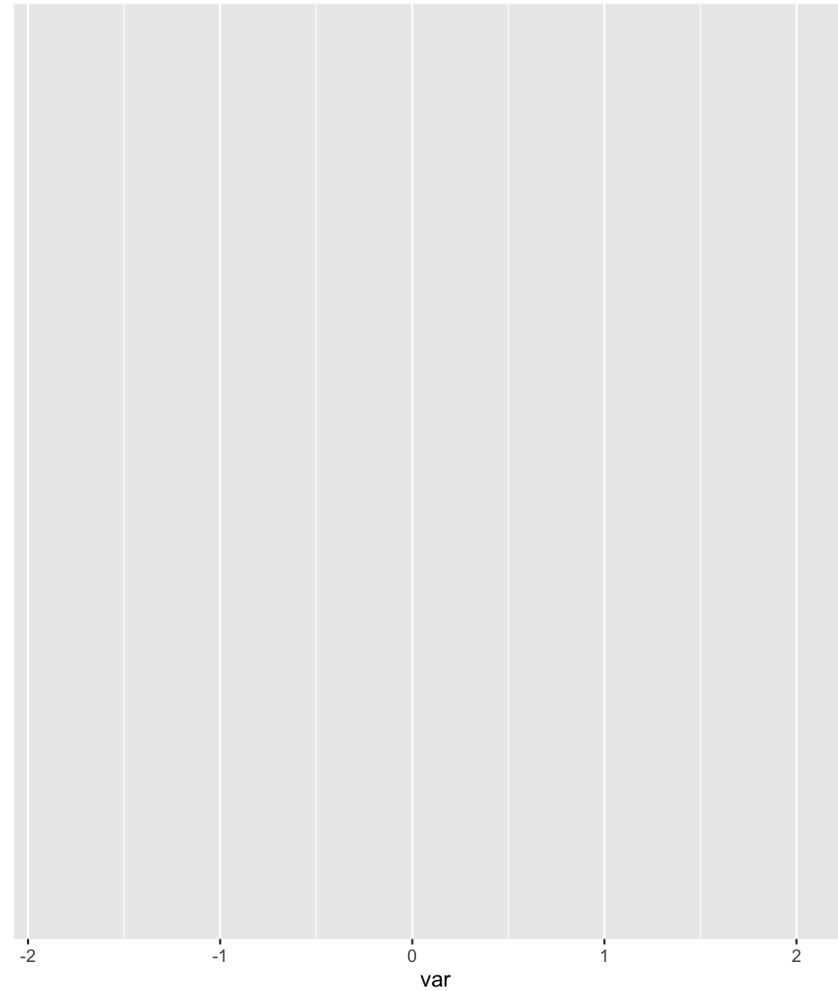
```

set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 20) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var)

```



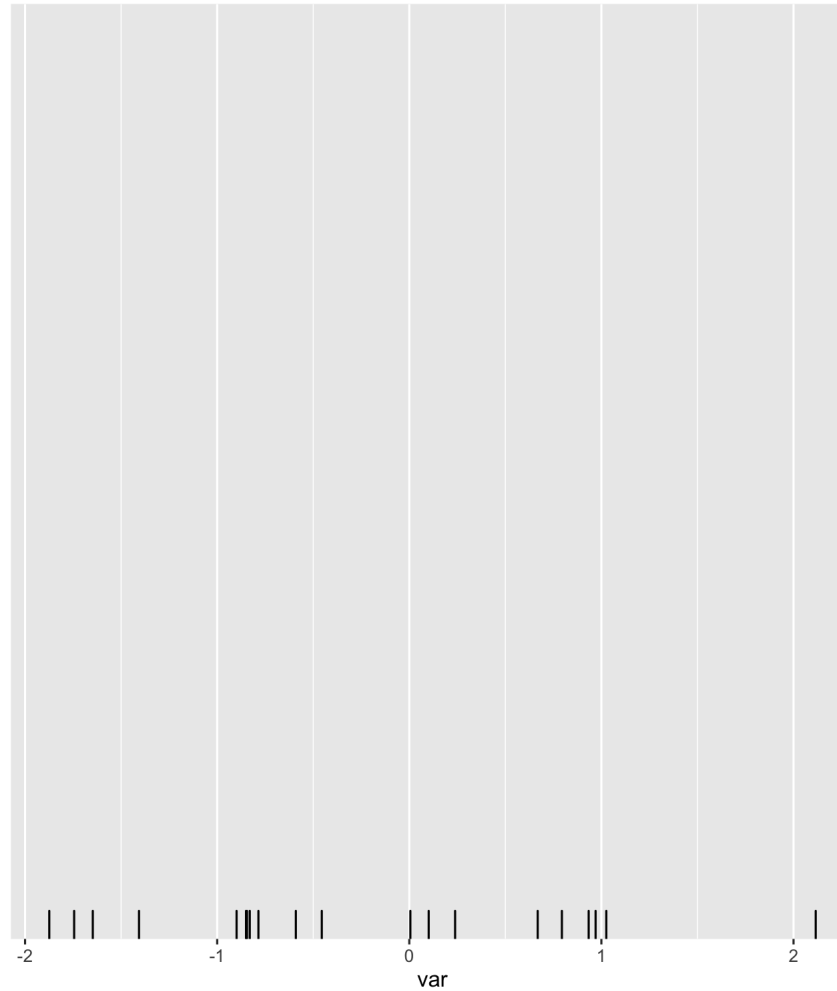
```

set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 20) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug()

```



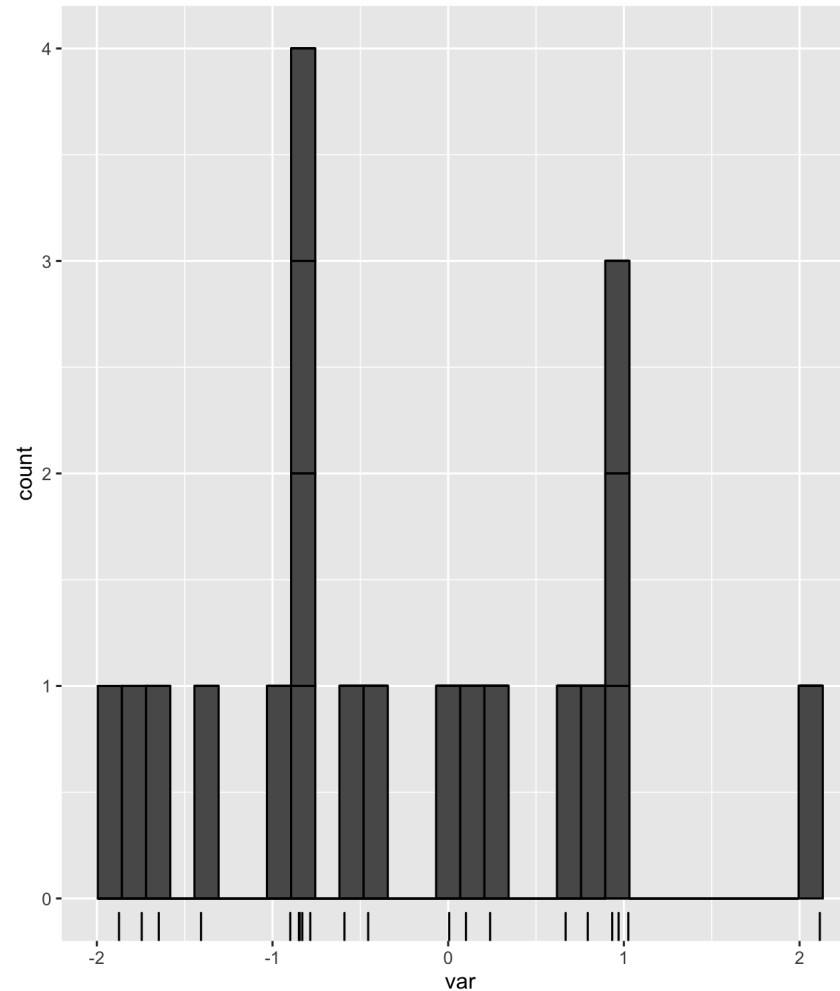
```

set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 20) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  geom_histogram(aes(group = index),
                 color = "black")

```



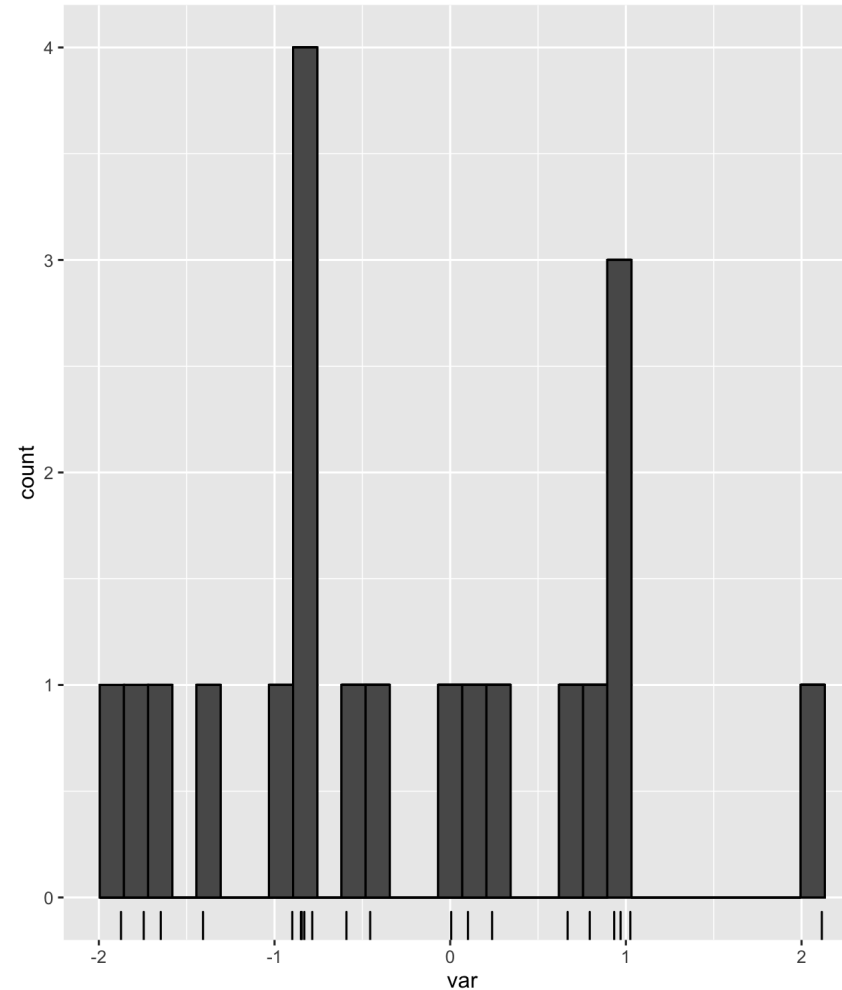
```

set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 20) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  geom_histogram(aes(group = index),
                 color = "black") +
  geom_histogram(color = "black")

```



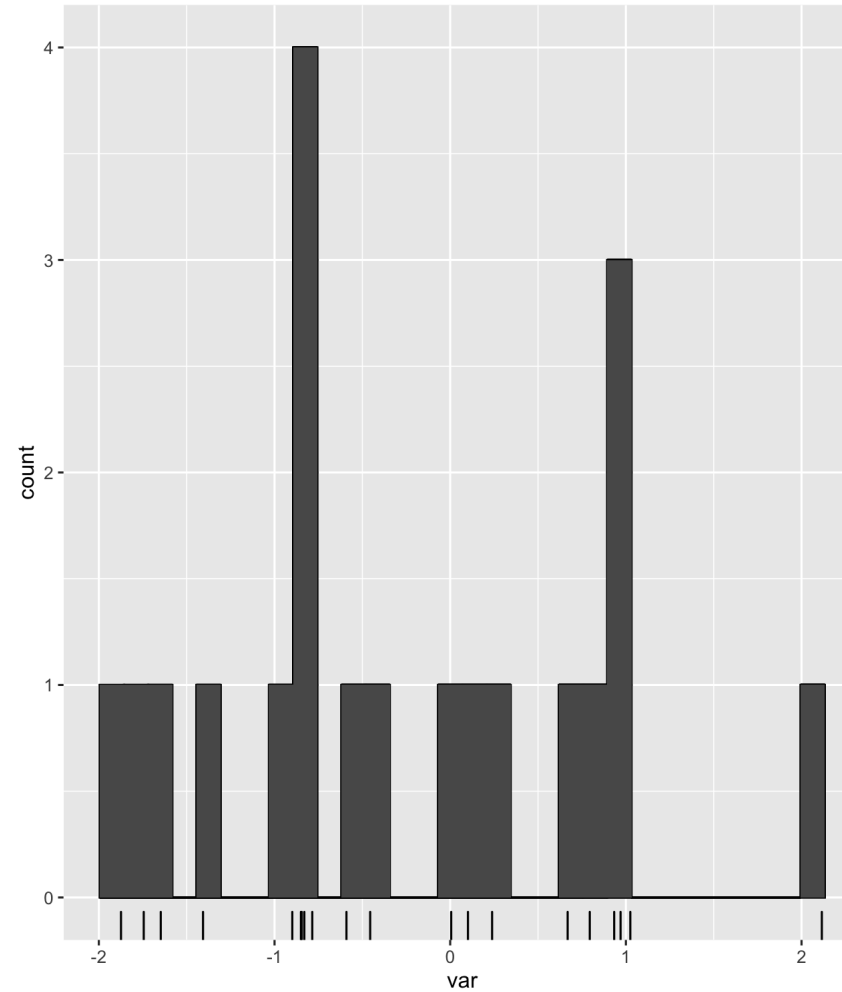
```

set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 20) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  geom_histogram(aes(group = index),
                 color = "black") +
  geom_histogram(color = "black") +
  geom_histogram()

```



```

set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 20) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  geom_histogram(aes(group = index),
                 color = "black") +
  geom_histogram(color = "black") +
  geom_histogram() ->
hist

```

```

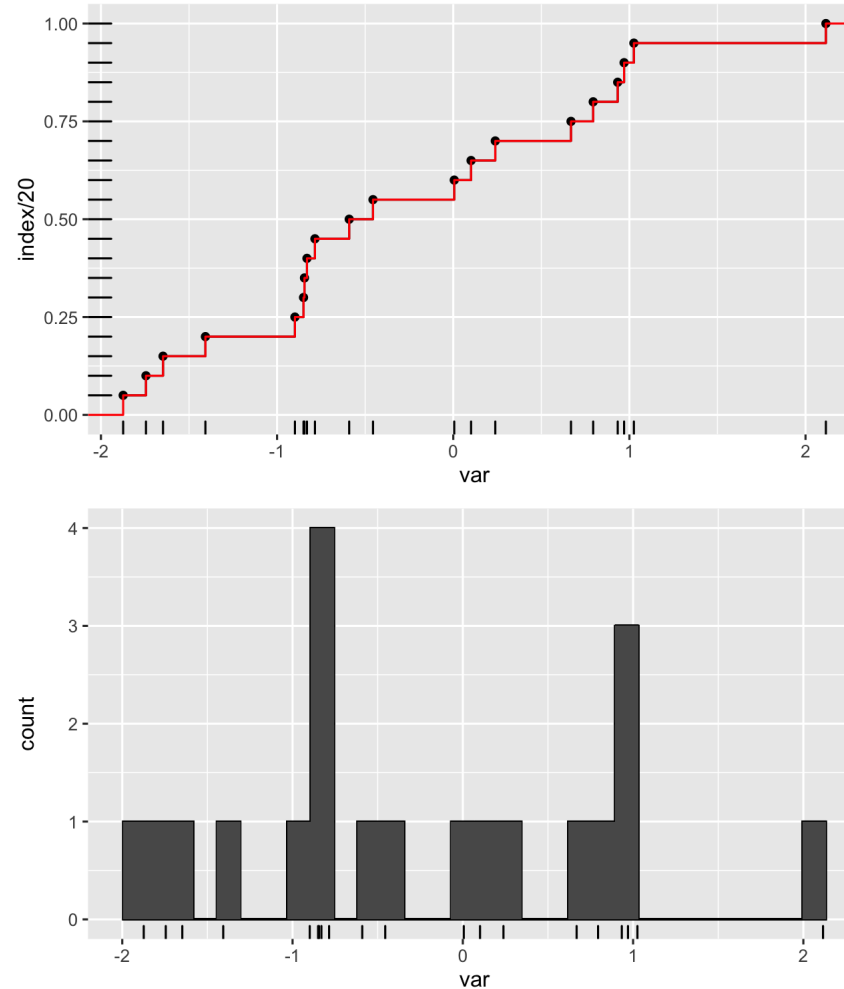
set.seed(780)
rnorm(20) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 20) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  geom_histogram(aes(group = index),
                 color = "black") +
  geom_histogram(color = "black") +
  geom_histogram() ->
hist

ecdf / hist

```



larger N


```
set.seed(780)
```

```
set.seed(780)
```

```
rnorm(200)
```

```
[1] -1.647531241 -0.898891586  0.005635507 -0.455534937  2.115503581
[6]  0.101123587 -1.873883654  0.969946060 -0.830382178  0.933596933
[11] -0.590622722 -1.406949386  0.238567023 -1.744497088 -0.849863390
[16]  1.025406106  0.794348885  0.668462054 -0.844740960 -0.785183840
[21]  1.837337685  0.442731604 -1.328259220 -0.281602466  1.080365048
[26] -0.575308566  0.139441906  0.334802755  0.033743973  1.319005208
[31]  0.053124840  3.255166528 -0.759656764 -1.538410325  1.748896588
[36]  0.613864628  1.380147538 -1.714593651 -1.069075429  0.375454550
[41] -0.677401509 -0.145892718 -0.483588641  1.406019850  1.017028354
[46] -0.098136653  0.690471475  0.540770949 -1.797859900 -0.754777810
[51] -0.853192425  1.129422400 -1.136723505 -1.434601881 -1.798525534
[56]  1.377375661 -0.269621749  0.254726943  0.256149208  0.996431286
[61] -0.271595632 -0.129531145  0.442837585 -1.423907600  0.260302095
[66]  0.932503395 -0.210942501  0.284693554 -0.743559167  0.145353392
[71] -1.218458641  1.342264819  0.611150523  0.321078553  1.049817170
[76] -0.517772666  0.463814953 -0.742677180  0.064251896 -0.298629008
[81] -1.008338844  1.243924887  0.333020599  2.430218989 -0.303526605
[86] -1.407729838  0.642826539 -0.531465832 -0.222828484  0.741715116
[91] -0.651011411 -0.149275029  1.235855523  1.205554765  1.021957027
[96]  1.805922364 -0.418401576  2.485920624 -0.687232603  0.712129802
[101]  0.392927948 -0.575877081  1.339105597 -0.295813269 -0.270682869
[106] -0.557272148 -1.476393319 -1.146370024  0.850123631 -0.964441072
[111] -0.376748668 -0.063408980  0.468356308 -0.808813077 -0.490751373
[116]  1.650534747 -0.274971200 -0.003575037  0.202407017  0.834817358
[121] -0.586922329 -1.155341496  0.119423556 -0.487350325 -0.305419643
[126]  0.922676005  1.201417234  0.061497497 -0.962727937 -0.009631657
[131] -0.192959006  0.715103542  1.435254106 -0.057136109 -1.336144518
[136]  1.370861804  0.981686393  0.941868149  1.453895961 -1.299029811
[141]  1.352799345  2.050433984  0.277279693 -0.964045765  1.006492599
[146] -0.450460771  0.545860105 -0.560415140  1.052360111 -0.217578442
[151] -0.759602267 -0.344148261 -0.719310085  0.753827059  0.148336721
[156]  1.479936906  0.443617745  1.679034469  1.274943550 -0.382573366
[161] -1.041572335 -0.699654620  0.588367809 -0.893171131 -0.610888130
```

```
set.seed(780)
rnorm(200) %>%
  tibble(var = .)
```

```
# A tibble: 200 x 1
      var
  <dbl>
1 -1.65
2 -0.899
3  0.00564
4 -0.456
5  2.12
6  0.101
7 -1.87
8  0.970
9 -0.830
10 0.934
# ... with 190 more rows
```

```
set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var)
```

```
# A tibble: 200 x 1
  var
<dbl>
1 -2.81
2 -2.34
3 -1.88
4 -1.87
5 -1.80
6 -1.80
7 -1.74
8 -1.71
9 -1.65
10 -1.62
# ... with 190 more rows
```

```
set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n())
```

```
# A tibble: 200 x 2
  var index
  <dbl> <int>
1 -2.81     1
2 -2.34     2
3 -1.88     3
4 -1.87     4
5 -1.80     5
6 -1.80     6
7 -1.74     7
8 -1.71     8
9 -1.65     9
10 -1.62    10
# ... with 190 more rows
```

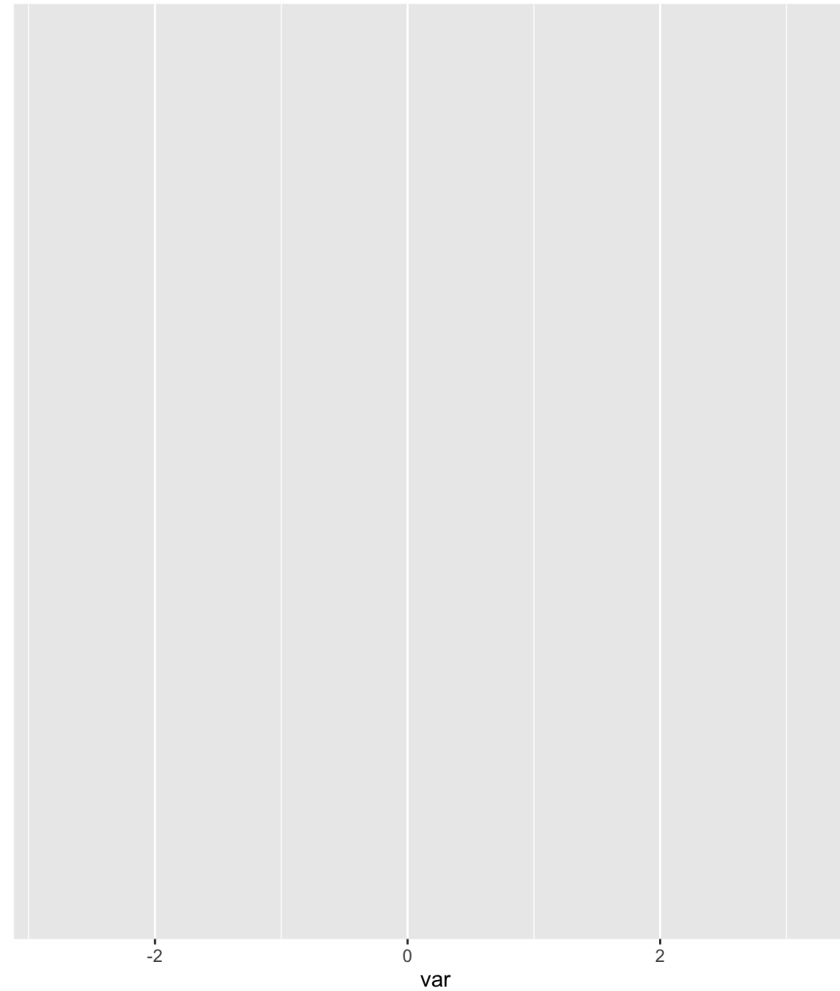
```
set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df
```

```
set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df)
```

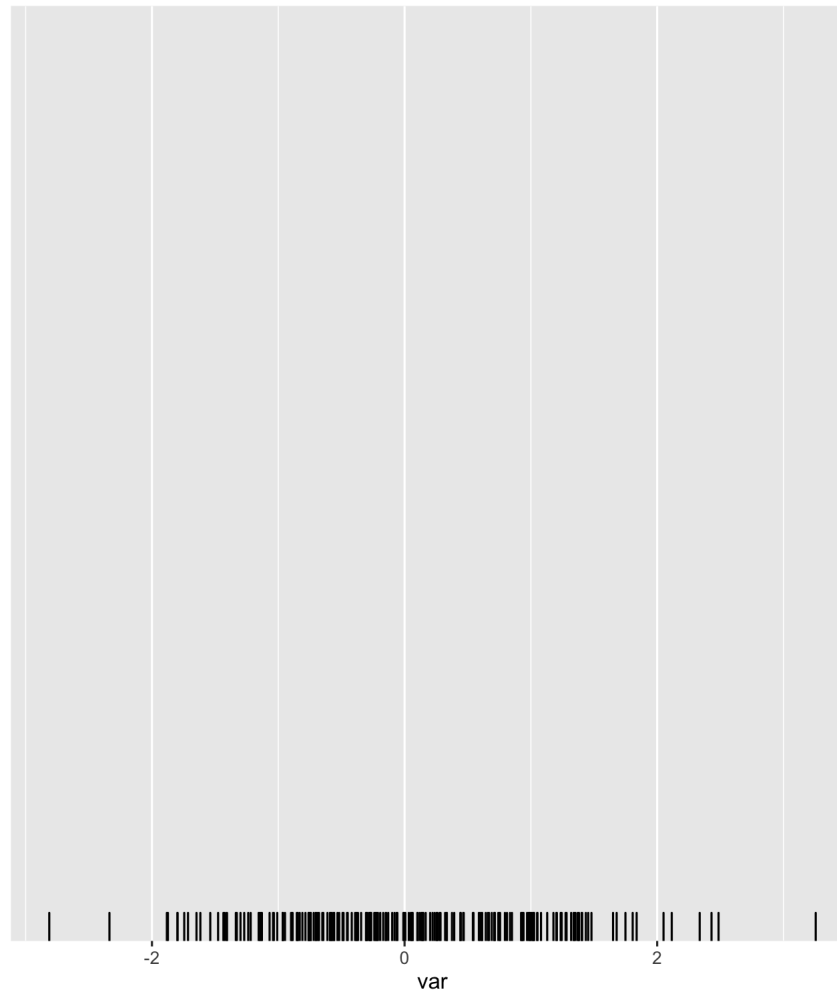
```
set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var)
```



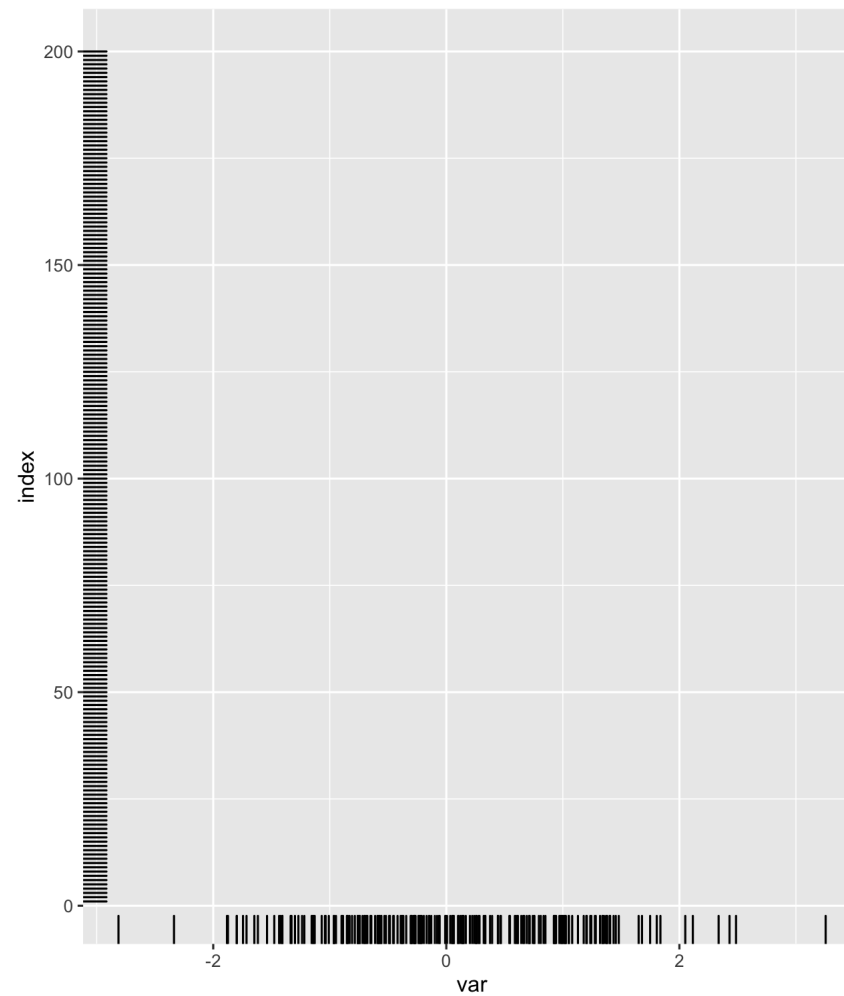

```
set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug()
```



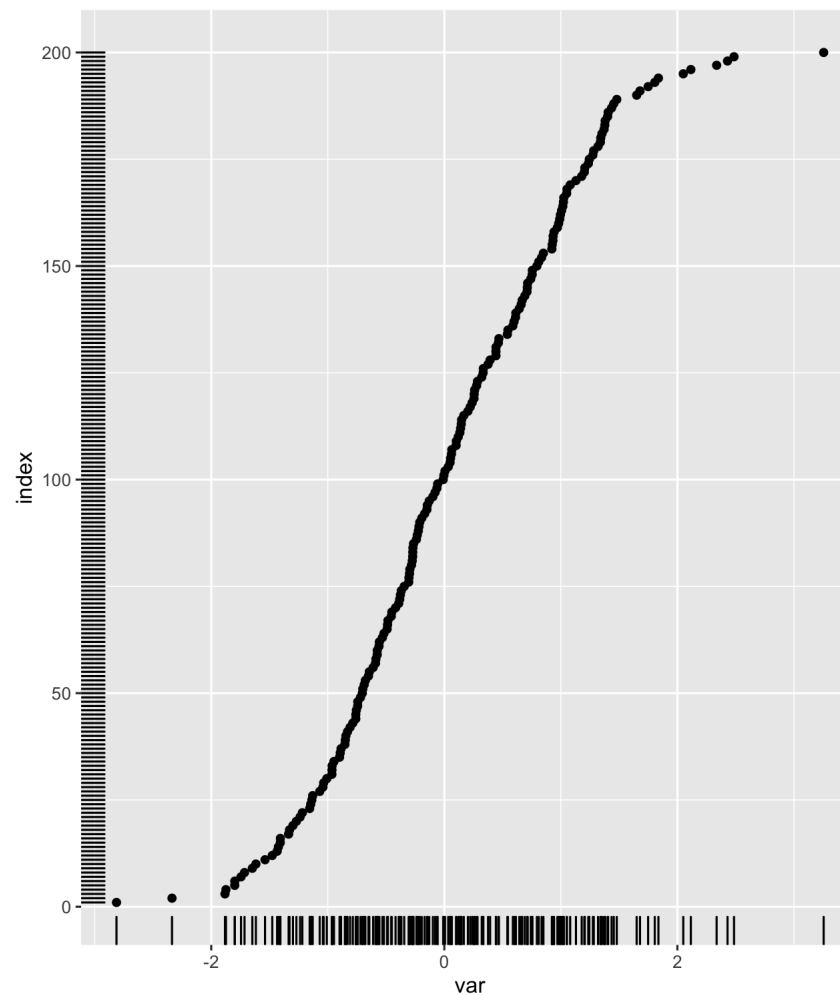
```
set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index)
```



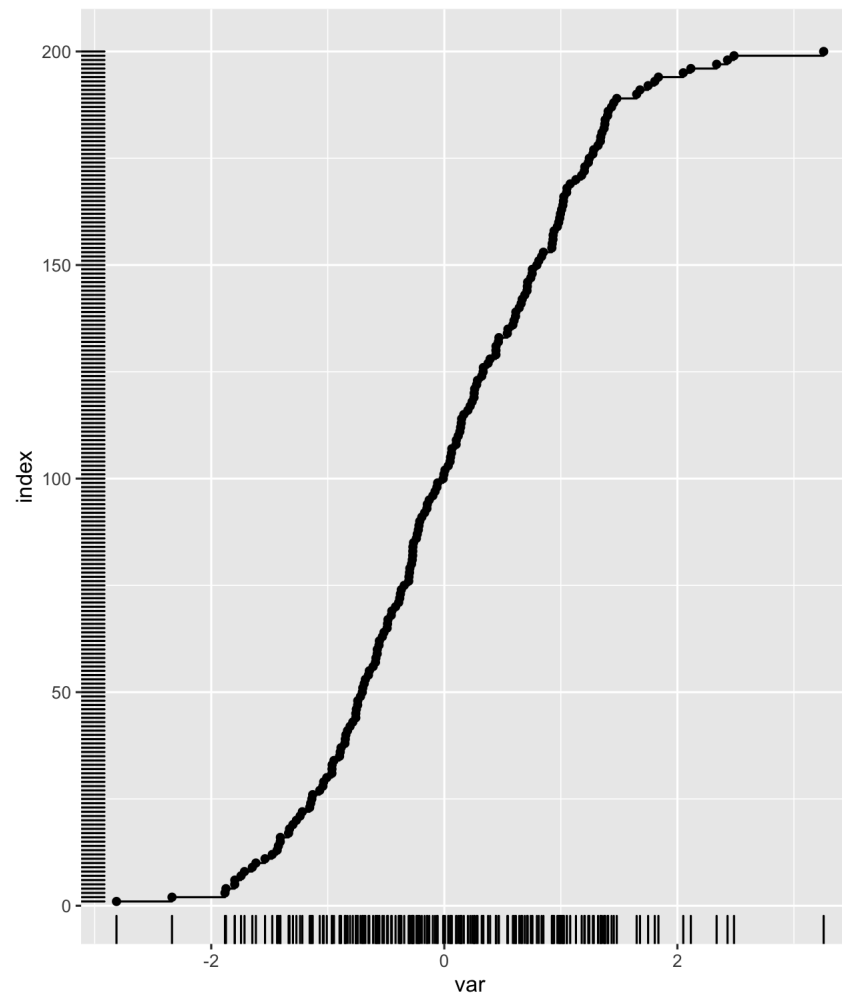
```
set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point()
```



```
set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step()
```

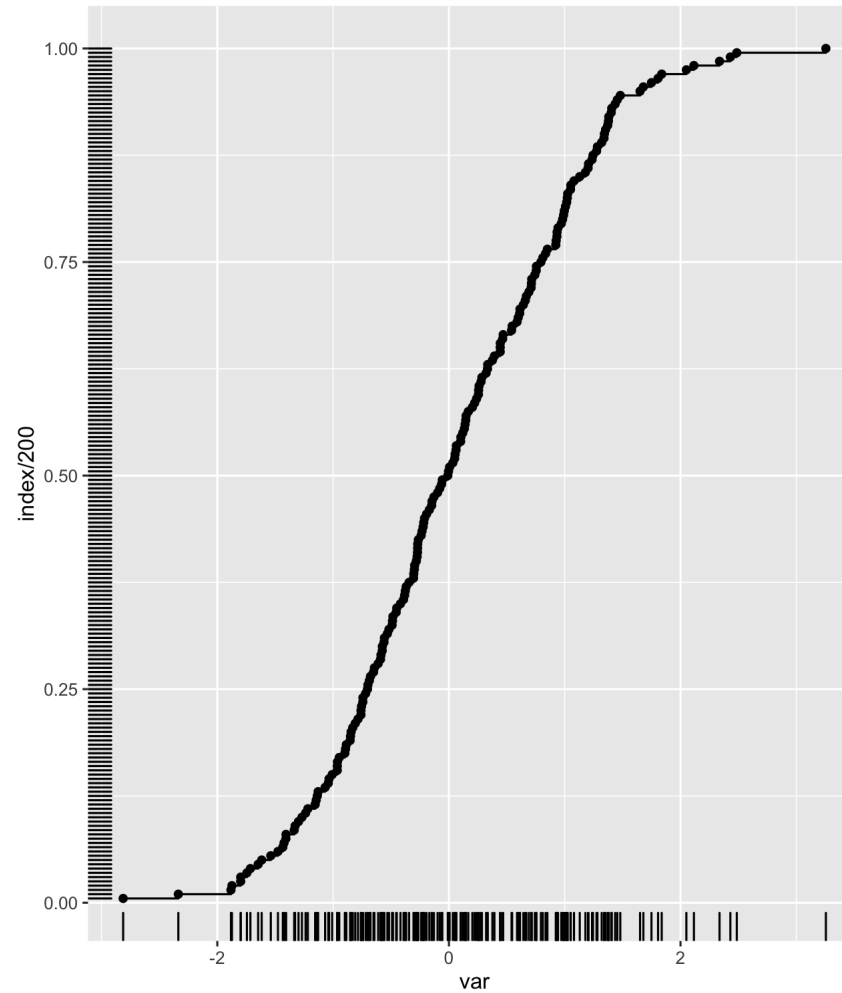


```

set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 200)

```

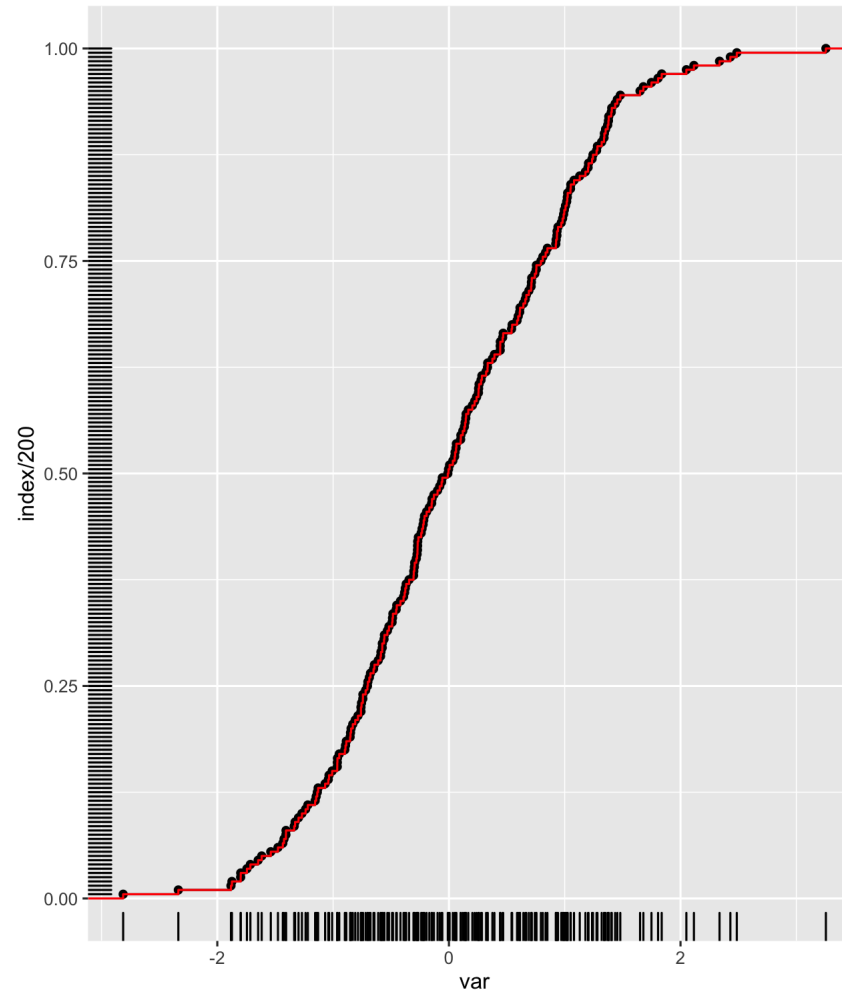


```

set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 200) +
  stat_ecdf(color = "red")

```



```
set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 200) +
  stat_ecdf(color = "red") ->
ecdf
```

```
set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 200) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df)
```



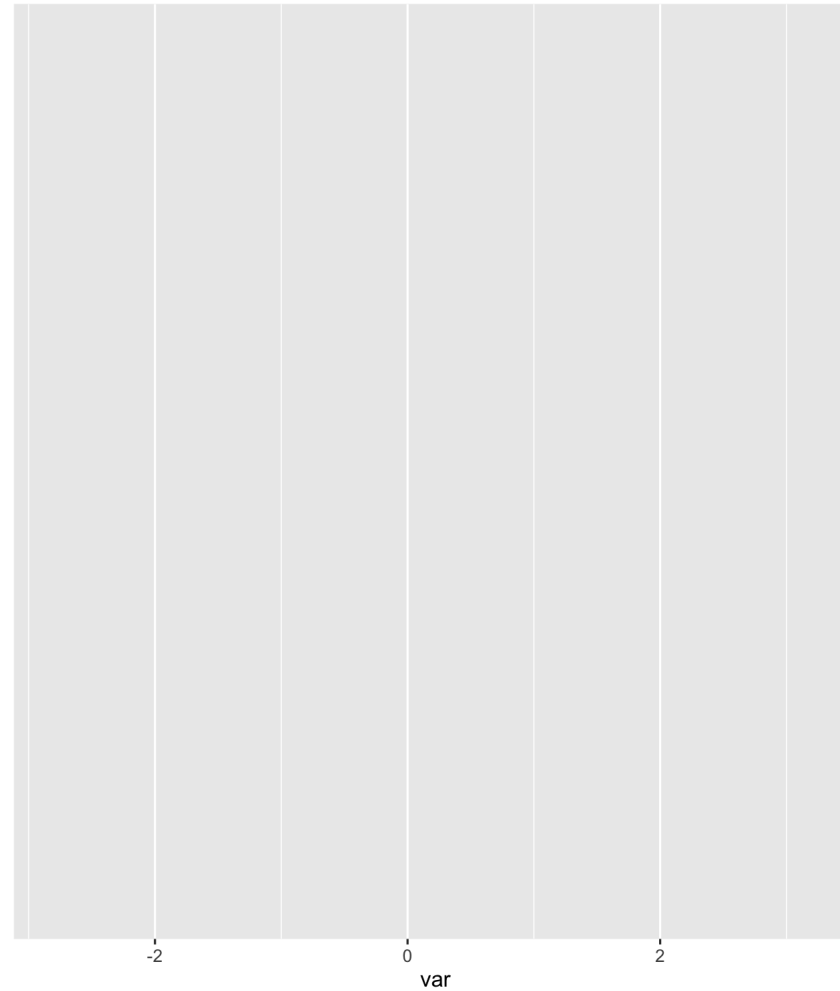
```

set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 200) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var)

```



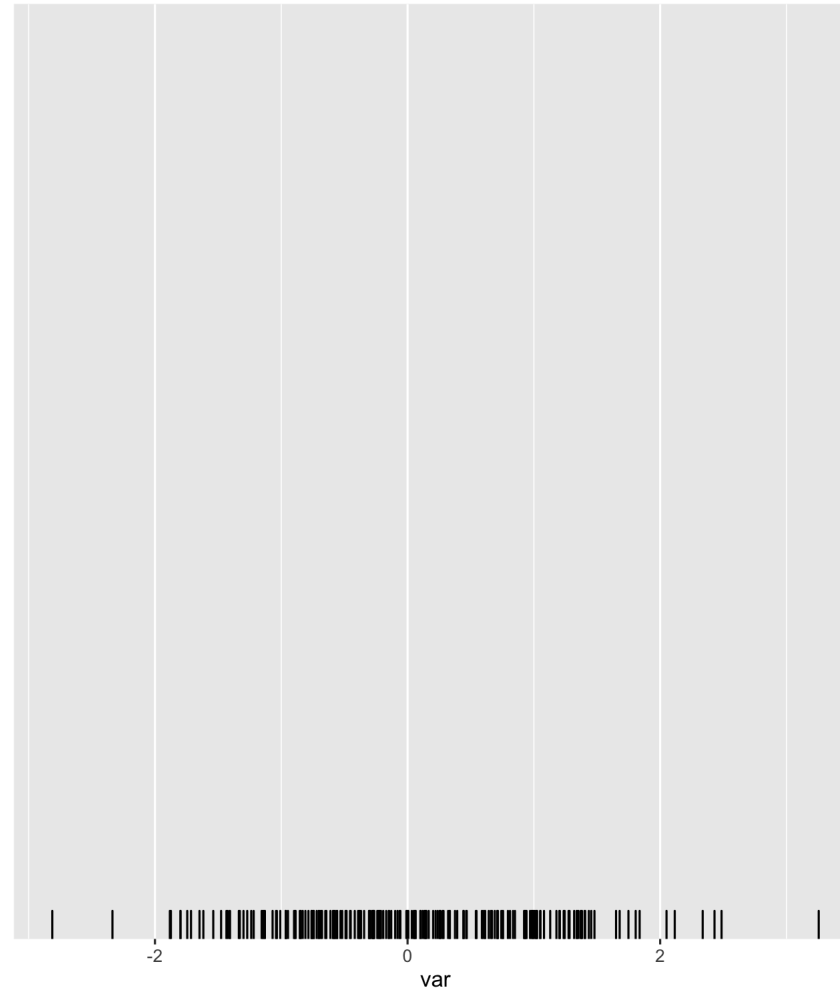
```

set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 200) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug()

```



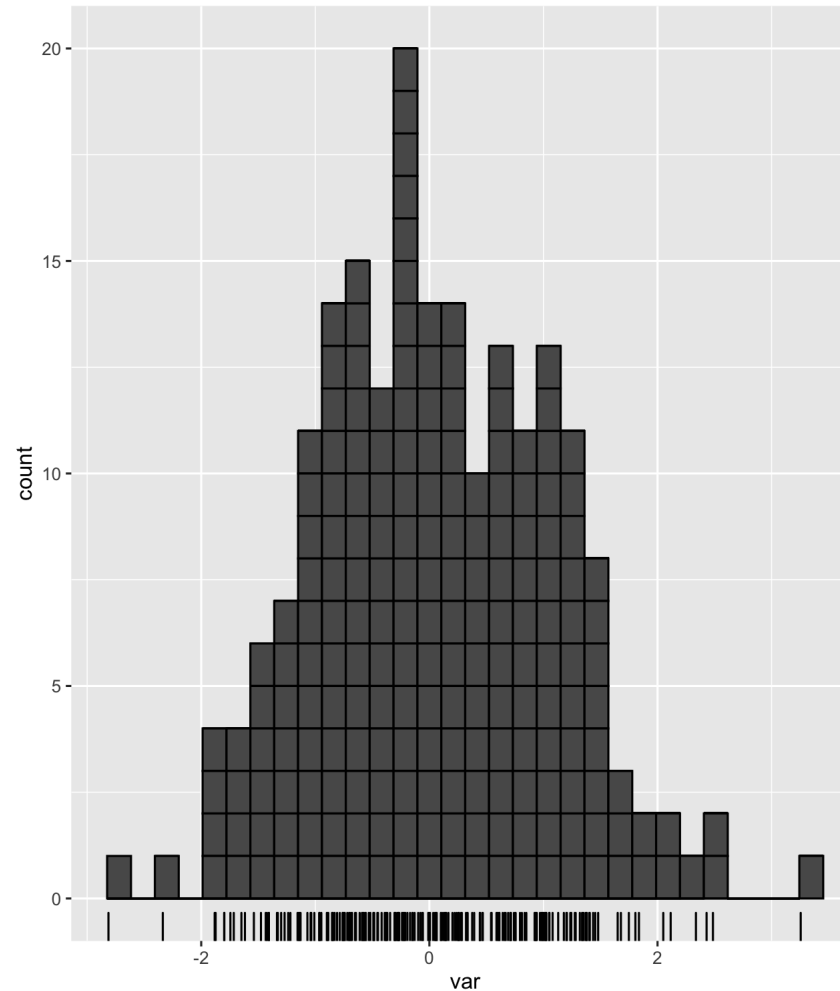
```

set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 200) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  geom_histogram(aes(group = index),
                 color = "black")

```



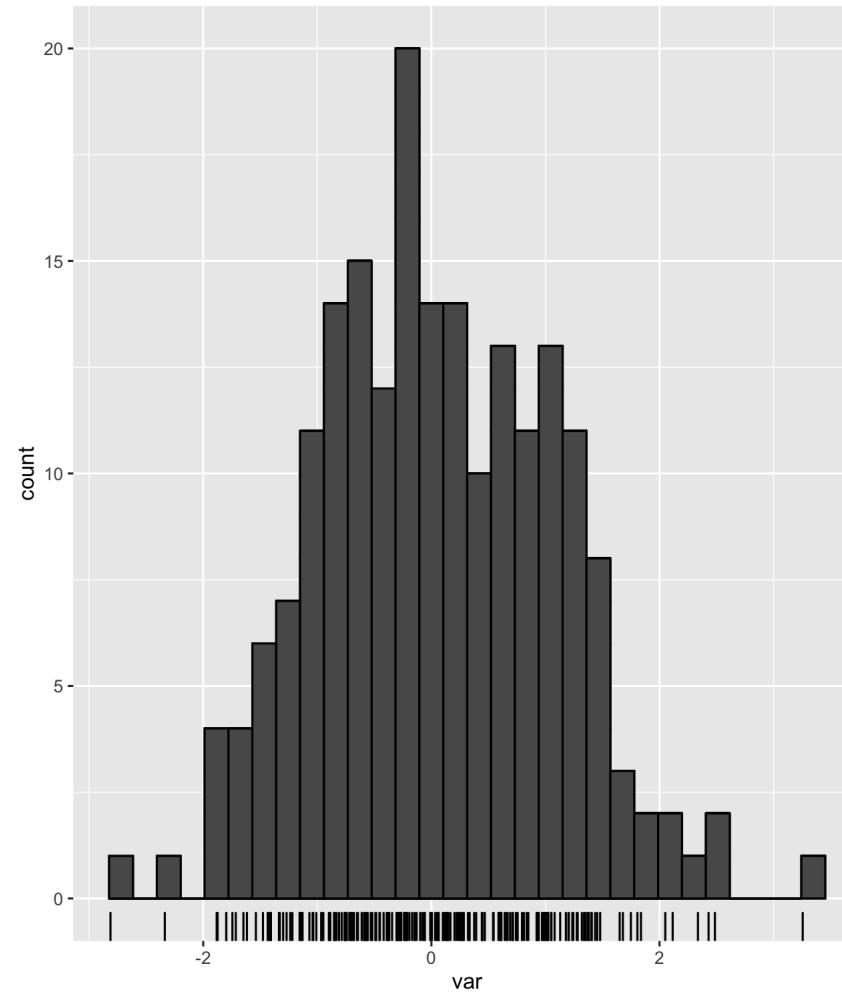
```

set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 200) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  geom_histogram(aes(group = index),
                 color = "black") +
  geom_histogram(color = "black")

```



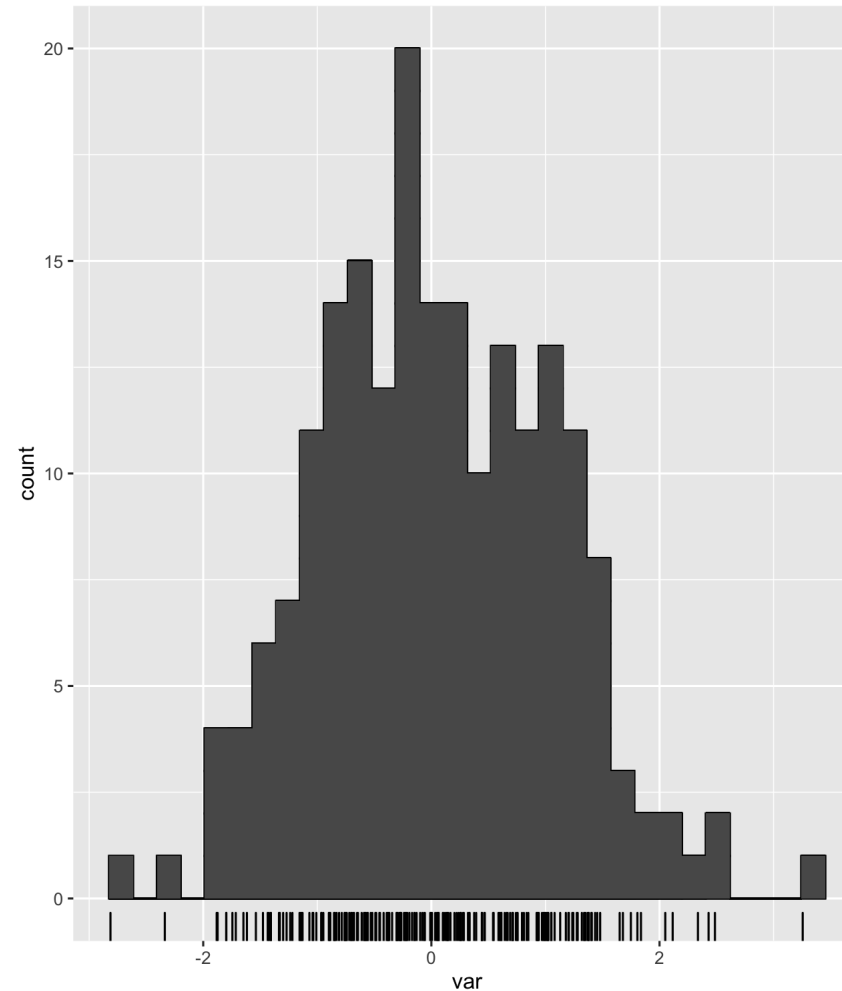
```

set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 200) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  geom_histogram(aes(group = index),
                 color = "black") +
  geom_histogram(color = "black") +
  geom_histogram()

```



```

set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 200) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  geom_histogram(aes(group = index),
                 color = "black") +
  geom_histogram(color = "black") +
  geom_histogram() ->
hist

```

```

set.seed(780)
rnorm(200) %>%
  tibble(var = .) %>%
  arrange(var) %>%
  mutate(index = 1:n()) ->
df

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  aes(y = index) +
  geom_point() +
  geom_step() +
  aes(y = index / 200) +
  stat_ecdf(color = "red") ->
ecdf

ggplot(data = df) +
  aes(x = var) +
  geom_rug() +
  geom_histogram(aes(group = index),
                 color = "black") +
  geom_histogram(color = "black") +
  geom_histogram() ->
hist

ecdf / hist

```

