Final Project

Yena Kim

2021 12 17

What question(s) am I interested in?

Why do international bureaucrats influence some recipient performance in World Bank projects but not others?

What are my independent and dependent variables? How do I plan to measure the variables?

My dependent variables will be staff-level variables, namely the influence of World Bank bureaucrats. I will employ the replication data of Heinzel and Liese (2021) to measure staff influence. They used three variables to measure international bureaucrats influence on recipient performance in World Bank projects: Staff Supervision, Staff Leniency, Staff Experience.

My independent variables is the preferences of U.S. over recipients (Affinity with the US). I will measure the ideal point distances from the US to operationalize the preferences of US. over recipients. Ideal point distances between the U.S. and recipients show how much the U.S. and recipients have different ideas in terms of their foreign policies as demonstrated in UN voting. This variable can also reflect how much the U.S. favor recipients.

Which dataset(s) will you use?

I will employ the replication data of Heinzel and Liese (2021).

```
load("C:/Rworkingdirectory/2021_Fall_R/811_R/Staff_and_Affinity.RData")
```

How would I visually represent them?

I plan to create three descriptive graphs.

1. The summary statistics of independent and dependent variables

I will highlight 'ave_ttl_super', 'ave_ttl_disbborr', 'log_meanexp' variables for my dependent variables and 'unga_us_imp' variable for my independent variable from the dataset labeled here 'Staff_ccode'. In the summary table, missing values are listed as "Unknown" in the table.

```
library(gtsummary)
```

```
## Warning: package 'gtsummary' was built under R version 4.1.2
```

```
sum_dep <- Staff_ccode %>% select(ave_ttl_super, ave_ttl_disbborr, log_meanexp)
sum_dep %>% tbl_summary(statistic = list(all_continuous() ~ "{mean} ({sd})"))
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**N = 3,962**
ave_ttl_super	4.36 (0.91)
Unknown	1,720
ave_ttl_disbborr	0.25 (0.09)
Unknown	1,910
log_meanexp	0.27 (0.46)
Unknown	1

```
sum_indep <- Staff_ccode %>% select(unga_us_imp)
sum_indep %>% tbl_summary(statistic = list(all_continuous() ~ "{mean} ({sd})"))
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**N = 3,962**
unga_us_imp	3.26 (0.83)
Unknown	251

For 'ave_ttl_super', the mean is 4.36, and the standard deviation is 0.91.

As for 'ave_ttl_disbborr', the mean is 0.25, and the standard deviation is 0.09.

'log_meanexp' variables has the mean of 0.27 and the standard deviation of 0.46.

Concerning the independent variable 'unga_us_imp', the mean is 3.26, and the standard deviation is 0.83.

2. The distribution of depedent and independent variables

A boxplot is a standardized way of displaying the distribution of data based on a five number summary ("minimum", first quartile (Q1), median, third quartile (Q3), and "maximum").

```
library("here")
```

```
## Warning: package 'here' was built under R version 4.1.1
```

```
## here() starts at C:/Rworkingdirectory/2021_Fall_R/811_R
```

```
library("tidyverse")
```

```
## Warning: package 'tidyverse' was built under R version 4.1.1
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.0.2     v forcats 0.5.1
```

```
## Warning: package 'tibble' was built under R version 4.1.1
```

```
## Warning: package 'tidyr' was built under R version 4.1.1
```

```
## Warning: package 'readr' was built under R version 4.1.1
```

```
## Warning: package 'dplyr' was built under R version 4.1.1
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library("tidylog")
```

```
## Warning: package 'tidylog' was built under R version 4.1.1
```

```
##
## Attaching package: 'tidylog'
```

```
## The following objects are masked from 'package:dplyr':
##
##     add_count, add_tally, anti_join, count, distinct, distinct_all,
##     distinct_at, distinct_if, filter, filter_all, filter_at, filter_if,
##     full_join, group_by, group_by_all, group_by_at, group_by_if,
##     inner_join, left_join, mutate, mutate_all, mutate_at, mutate_if,
##     relocate, rename, rename_all, rename_at, rename_if, rename_with,
##     right_join, sample_frac, sample_n, select, select_all, select_at,
##     select_if, semi_join, slice, slice_head, slice_max, slice_min,
##     slice_sample, slice_tail, summarise, summarise_all, summarise_at,
##     summarise_if, summarize, summarize_all, summarize_at, summarize_if,
##     tally, top_frac, top_n, transmute, transmute_all, transmute_at,
##     transmute_if, ungroup
```

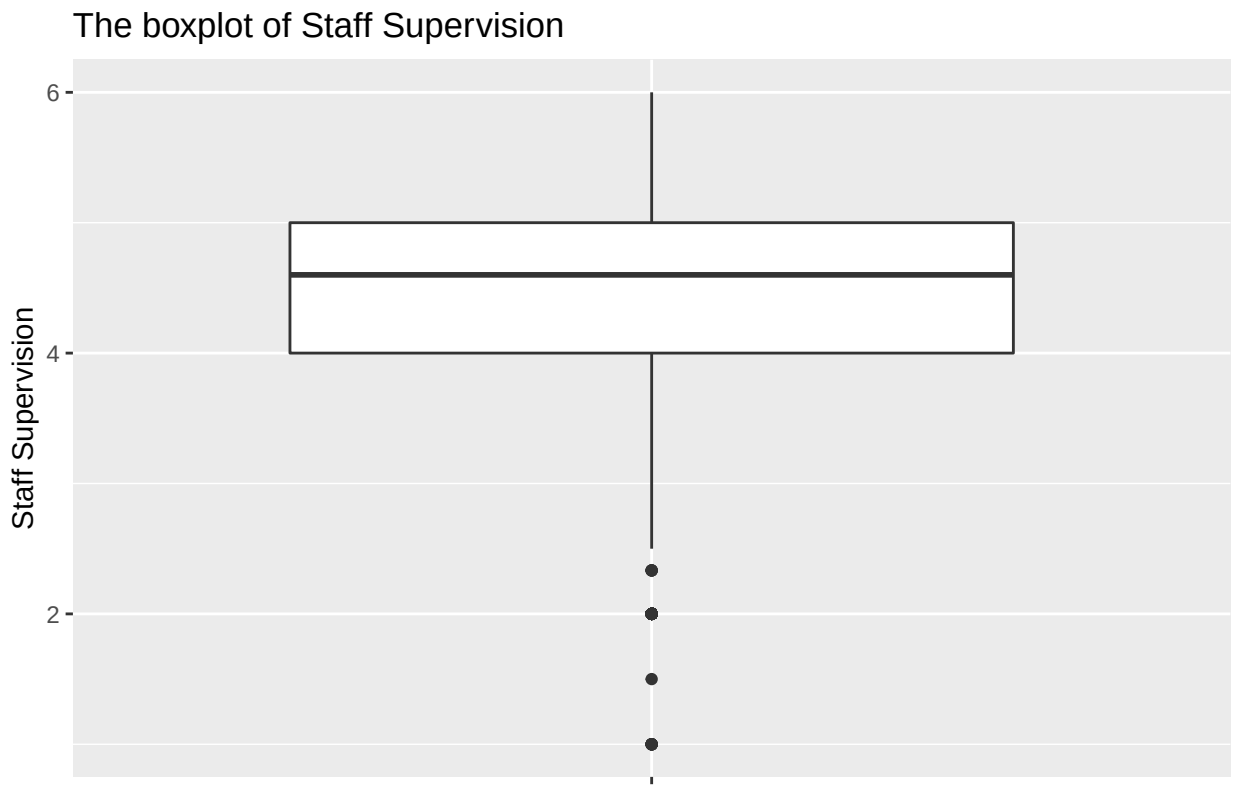
```
## The following objects are masked from 'package:tidyr':
##
##     drop_na, fill, gather, pivot_longer, pivot_wider, replace_na,
##     spread, uncount
```

```
## The following objects are masked from 'package:gtsummary':
##
##     mutate, select
```

```
## The following object is masked from 'package:stats':
##
##     filter
```

```
library("ggplot2")

ggplot(Staff_ccode, aes(x = " ", y = ave_ttl_super))+
  geom_boxplot()+
  labs(title="The boxplot of Staff Supervision",
       x=" ", y = "Staff Supervision")
```

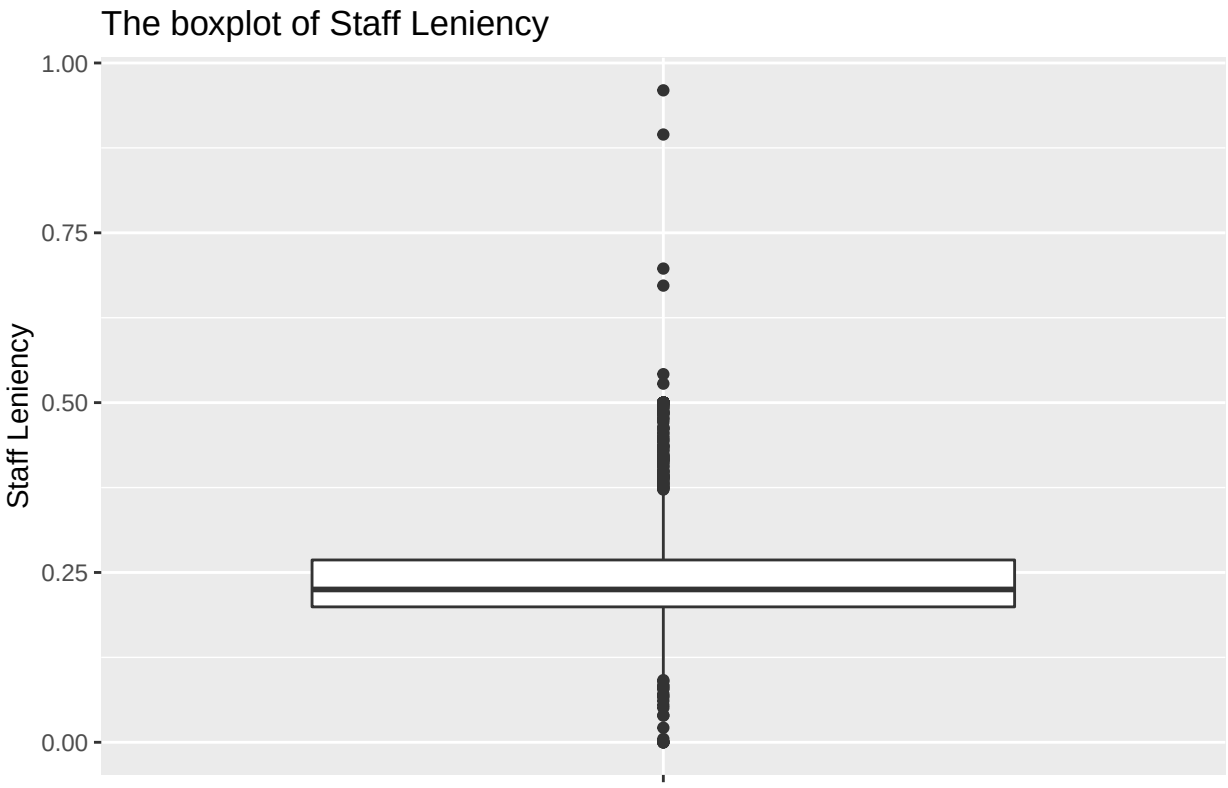
```
## Warning: Removed 1720 rows containing non-finite values (stat_boxplot).
```



The boxplot of Staff Supervision shows that the median of this variable is around 4.7.

```
ggplot(Staff_ccode, aes(x = " ", y = ave_ttl_disbborr))+
  geom_boxplot()+
  labs(title="The boxplot of Staff Leniency",
       x=" ", y = "Staff Leniency")
```

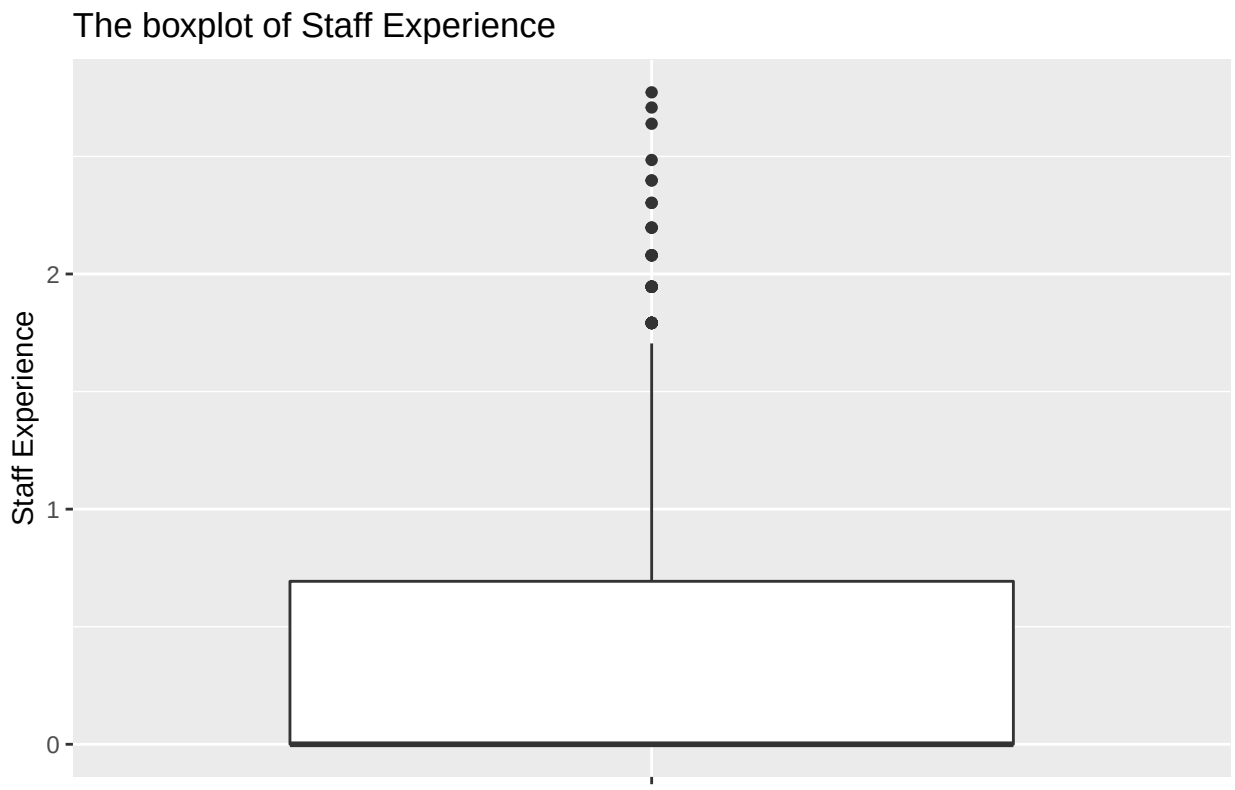
```
## Warning: Removed 1910 rows containing non-finite values (stat_boxplot).
```



The boxplot of Staff Leniency presents that the median is less than the mean of 0.25.

```
ggplot(Staff_ccode, aes(x = " ", y = log_meanexp))+
  geom_boxplot()+
  labs(title="The boxplot of Staff Experience",
       x=" ", y = "Staff Experience")
```

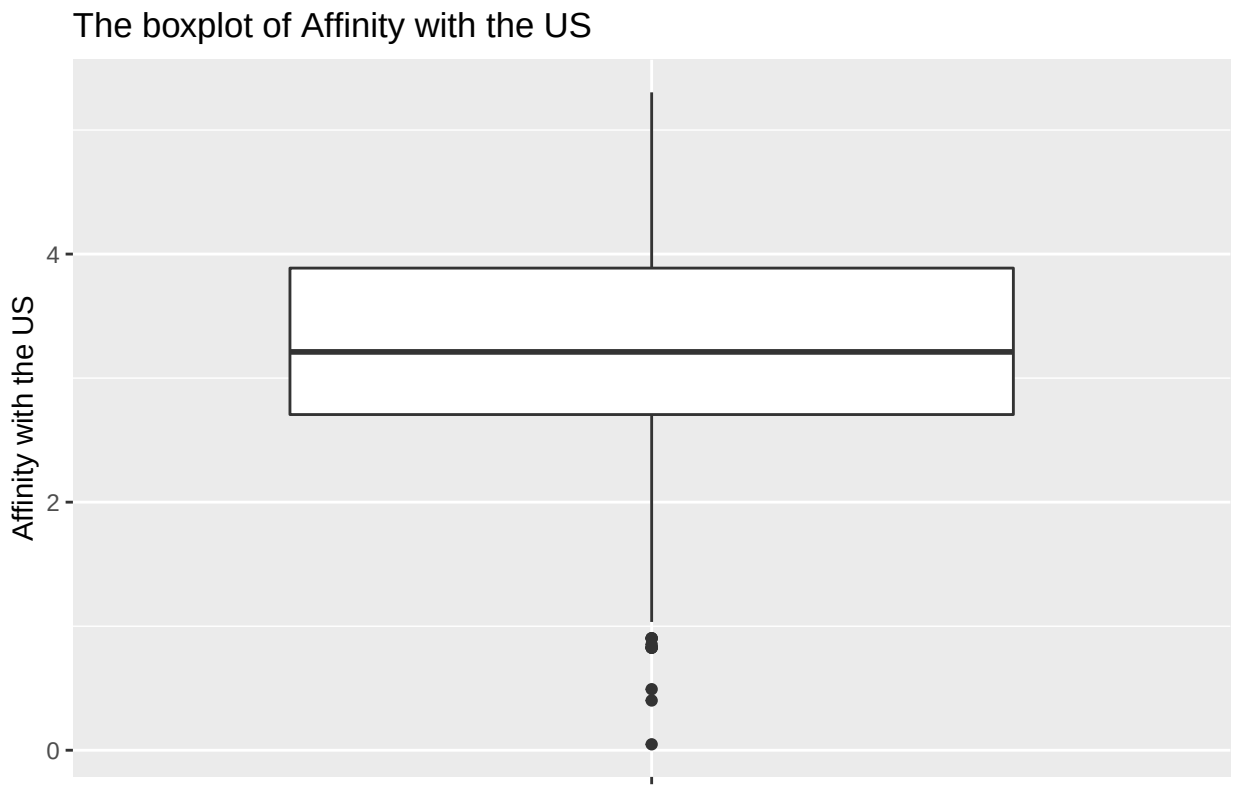
```
## Warning: Removed 1 rows containing non-finite values (stat_boxplot).
```



The boxplot of Staff Experience shows that median and minimum value are close to zero.

```
ggplot(Staff_ccode, aes(x = " ", y = unga_us_imp))+
  geom_boxplot()+
  labs(title="The boxplot of Affinity with the US",
       x=" ", y = "Affinity with the US")
```

```
## Warning: Removed 251 rows containing non-finite values (stat_boxplot).
```



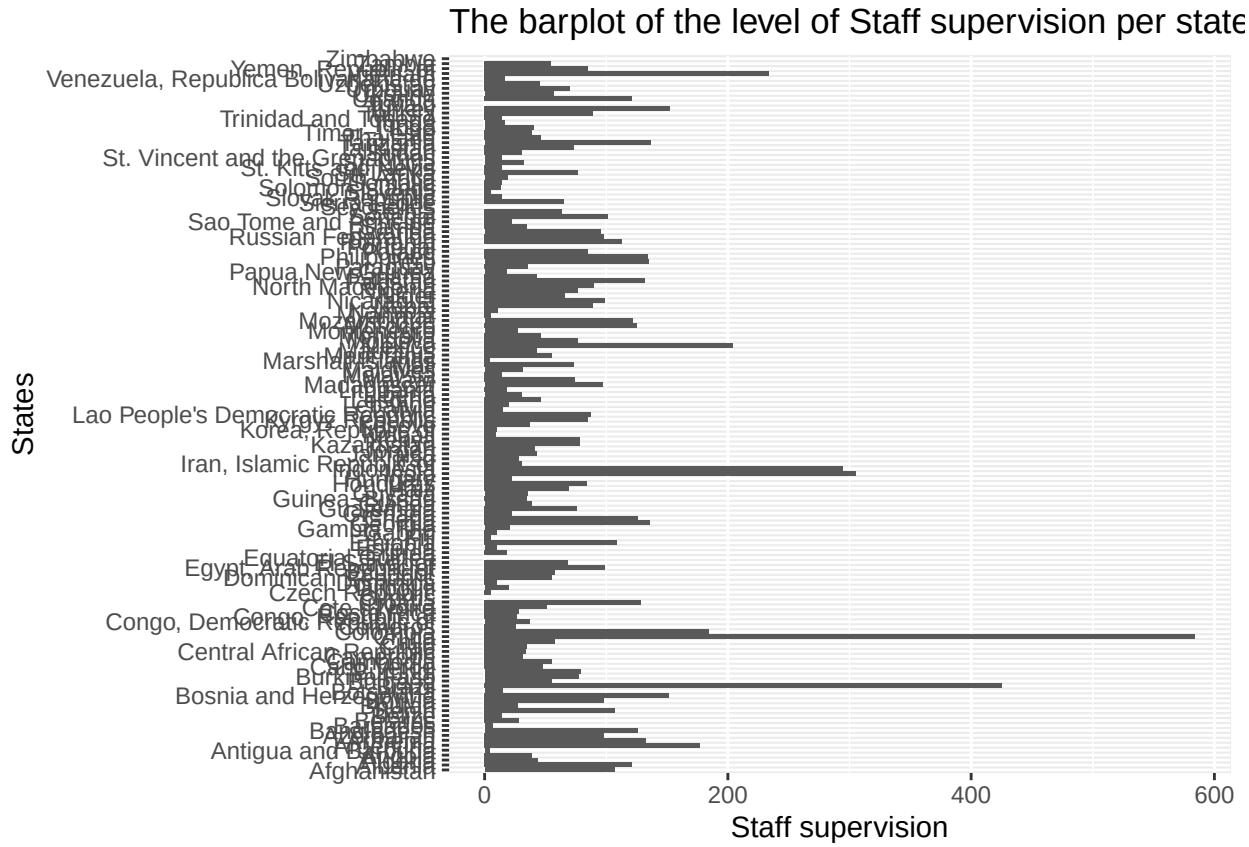
The boxplot of Affinity with the US presents that the median of this variable is around 3.2.

3. Variation of Staff influence grouped by countries

Three graph below presents the level of Staff supervision per state, the level of Staff leniency per state, and the level of Staff experience per state, respectively. It indicates there is a huge variation through country.

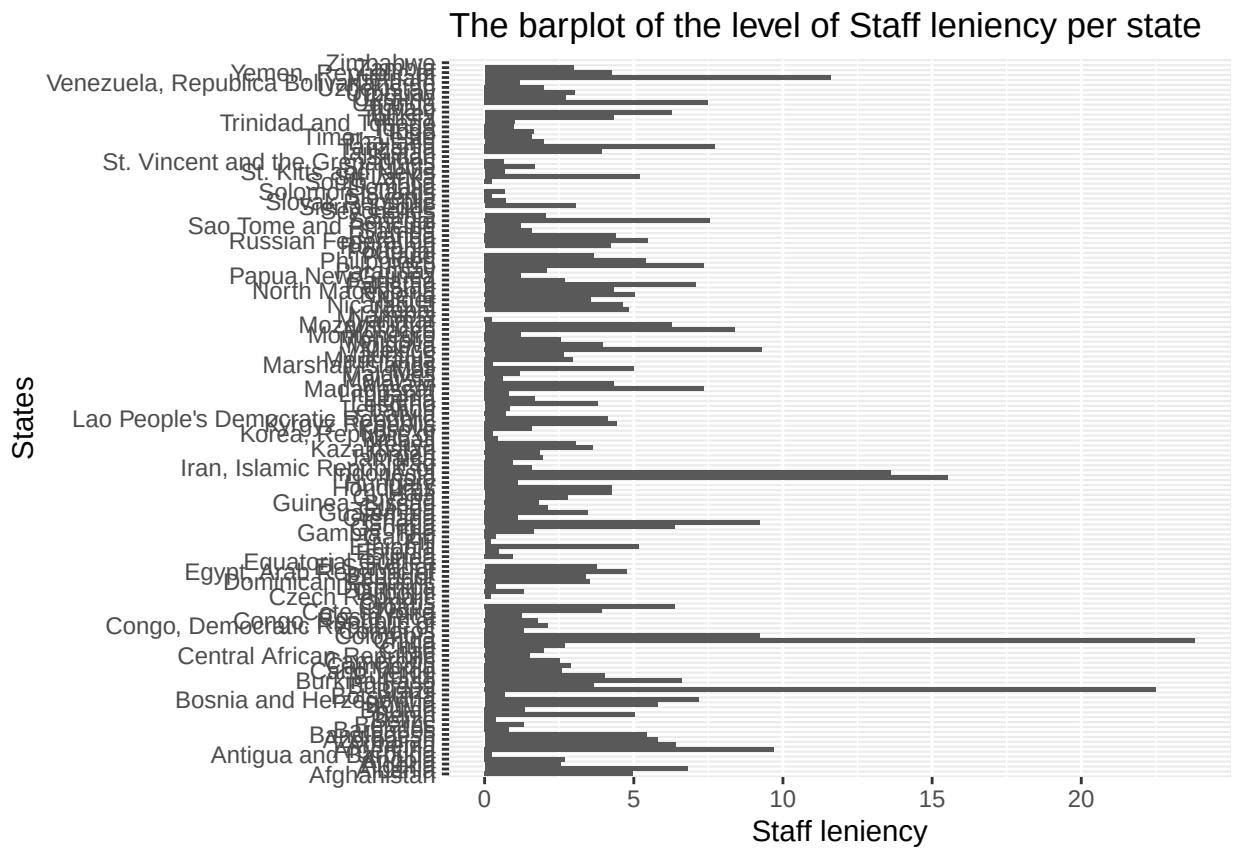
```
library(dplyr)

ggplot(Staff_ccode, aes(x = ave_ttl_super, y = countryname)) +
  geom_bar(stat="identity")+
  labs(title="The barplot of the level of Staff supervision per state",
       x= "Staff supervision", y = "States")
```

```
## Warning: Removed 1720 rows containing missing values (position_stack).
```



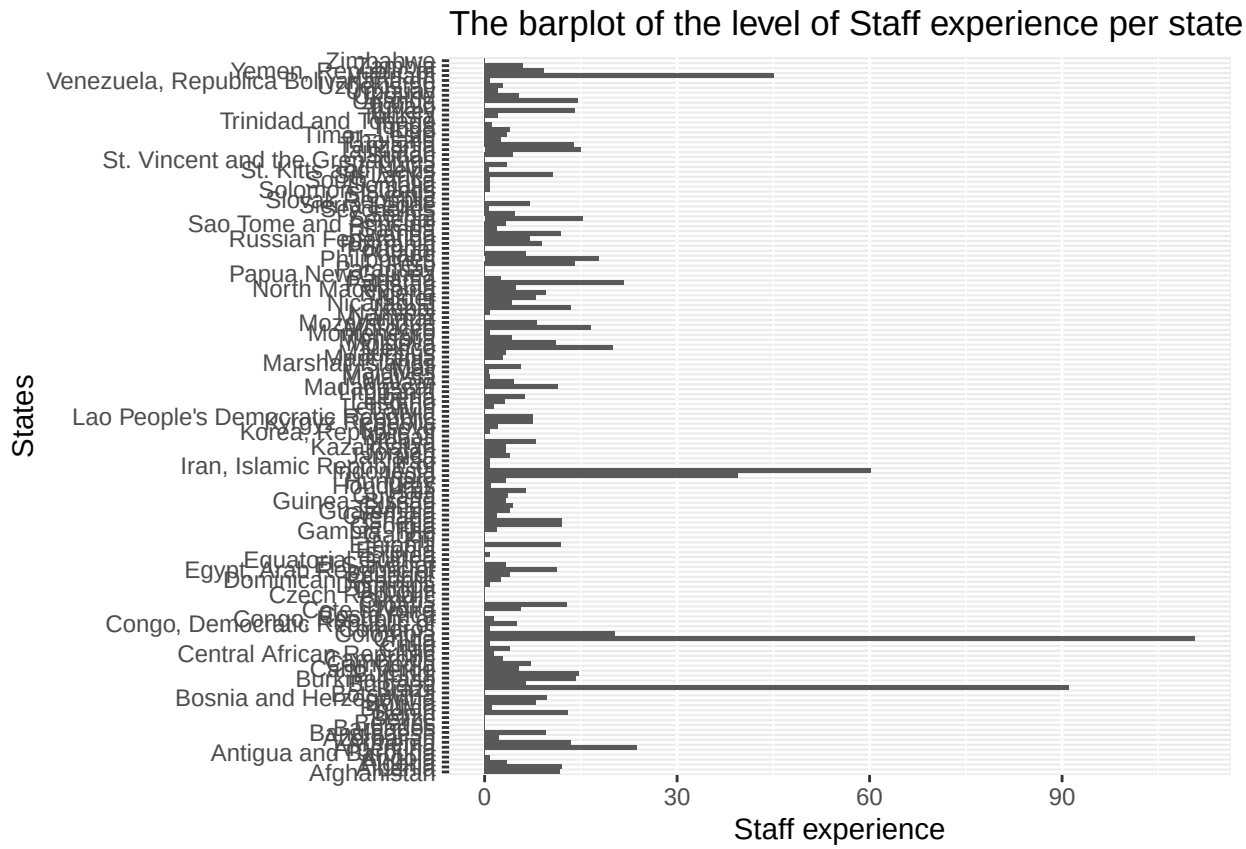

```
ggplot(Staff_ccode, aes(x = ave_ttl_disbborr, y = countryname)) +
  geom_bar(stat="identity")+
  labs(title="The barplot of the level of Staff leniency per state",
       x= "Staff leniency", y = "States")
```

```
## Warning: Removed 1910 rows containing missing values (position_stack).
```




```
ggplot(Staff_ccode, aes(x = log_meanexp, y = countryname)) +
  geom_bar(stat="identity")+
  labs(title="The barplot of the level of Staff experience per state",
       x= "Staff experience", y = "States")
```

```
## Warning: Removed 1 rows containing missing values (position_stack).
```

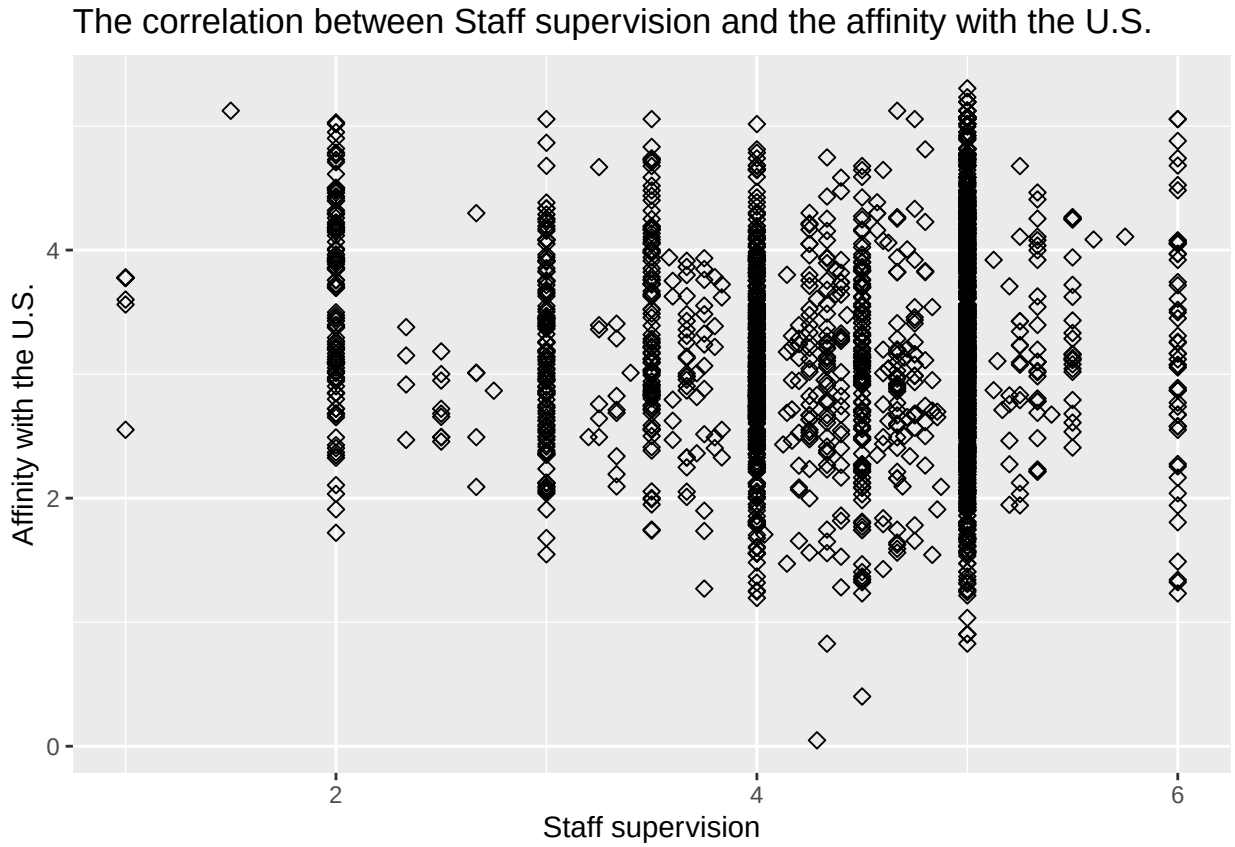


3. The correlation between Staff influence and the affinity with the U.S. using `ggplot2`.

Three graph below presents the correlation between the level of Staff supervision and the affinity with the U.S., the correlation between the level of Staff leniency and the affinity with the U.S., and the correlation between the level of Staff experience and the affinity with the U.S., respectively.

```
ggplot(Staff_ccode, aes(x=ave_ttl_super, y=unga_us_imp)) +
  geom_point(size=2, shape=23)+
   labs(title="The correlation between Staff supervision and the affinity with the U.S.",
       x="Staff supervision", y = "Affinity with the U.S.")
```

```
## Warning: Removed 1862 rows containing missing values (geom_point).
```




```
ggplot(Staff_ccode, aes(x=ave_ttl_disbborr, y=unga_us_imp)) +
  geom_point(size=2, shape=23)+
   labs(title="The correlation between Staff supervision and the affinity with the U.S.",
       x="Staff leniency", y = "Affinity with the U.S.")
```

```
## Warning: Removed 2040 rows containing missing values (geom_point).
```


```
ggplot(Staff_ccode, aes(x=log_meanexp, y=unga_us_imp)) +
  geom_point(size=2, shape=23)+
   labs(title="The correlation between Staff supervision and the affinity with the U.S.",
       x="Staff experience", y = "Affinity with the U.S.")
```

```
## Warning: Removed 252 rows containing missing values (geom_point).
```


The correlation between Staff supervision and the affinity with the U.S.
Affinity with the U.S.
0
2
4
0
1
2
Staff experience