Introduction to R

4.2 Plotting Anything

Example: Scatterplots
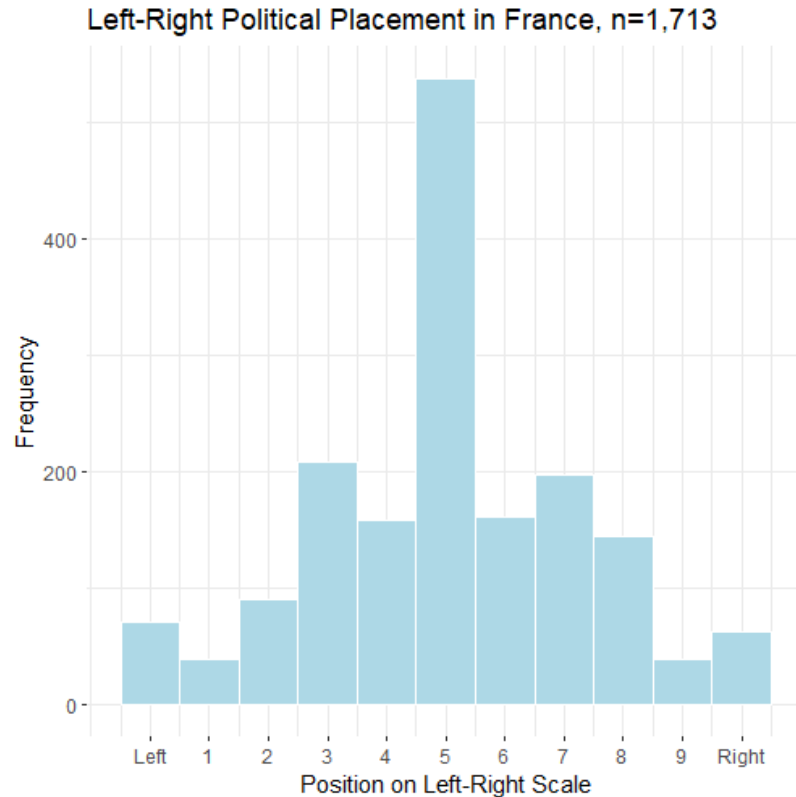
Lion Behrens, M.Sc.

UNIVERSITY OF MANNHEIM

University of Mannheim
Chair of Social Data Science and Methodology
Chair of Quantitative Methods in the Social
Sciences
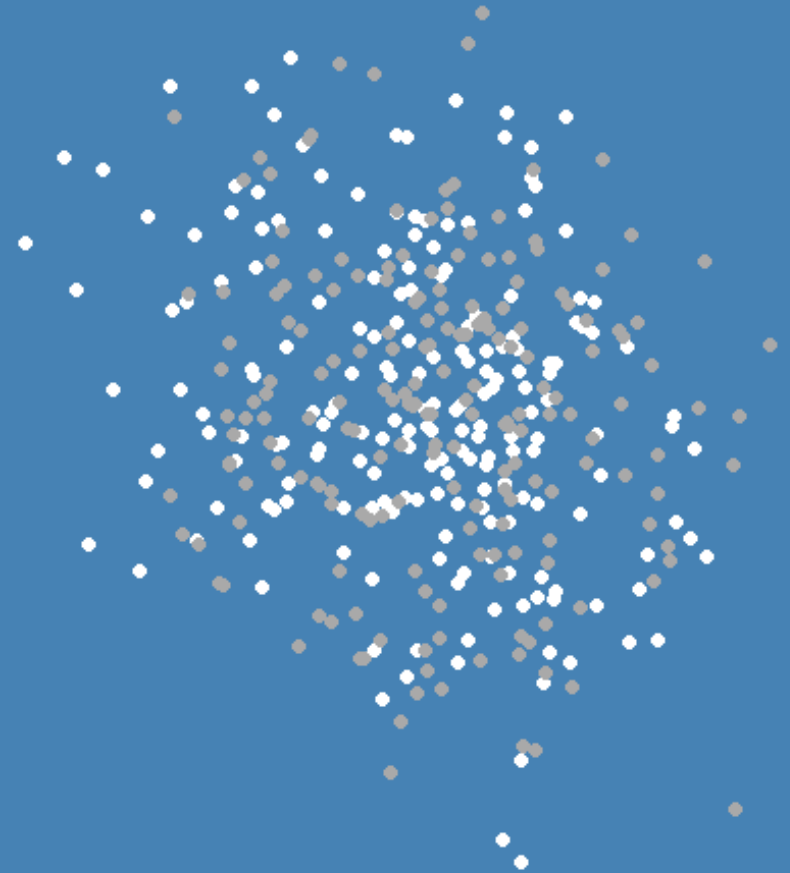
What we did so far

Our first plot: A Histogram

Left-Right Political Placement in France, n=1,713

```
ggplot(data = ess10,
       aes(x = left_right)) +
  geom_histogram(
    bins = 11, color = "white",
    fill = "lightblue") +
  scale_x_continuous(
    breaks = seq(0,10,1),
    labels = c("Left", seq(1,9,1), "Right")) +
  labs(x = "Position on Left-Right Scale",
       y = "Frequency",
       title = "Left-Right Political Placement
  theme_bw(base_size=15) +
  theme(panel.border = element_blank())
```

Steps for Plotting: Summary

A. Specify necessary layers

1. Data.
2. Aesthetics (*axis definitions*).
3. Geometric objects to use (*plot type*).

```
ggplot(data = ess10,
       aes(x = left_right)) +
  geom_histogram()
```

B. Customize your plot

1. Shape, color and size of geometric objects.
2. Fine-tune your x-scale and y-scale (axis ticks, ticks labels).
3. Label your axes, assign a title to your plot.
4. Choose a ggtheme.
5. Customize your theme.

Now: Plotting anything!

Steps for Plotting: Summary

A. Specify necessary layers

1. Data.
2. Aesthetics (*axis definitions*).
3. Geometric objects to use (*plot type*).

```
ggplot(data = ess10,
       aes(x = left_right)) +
  geom_histogram()
```

B. Customize your plot

1. Shape, color and size of geometric objects.
2. Fine-tune your x-scale and y-scale (axis ticks, ticks labels).
3. Label your axes, assign a title to your plot.
4. Choose a ggtheme.
5. Customize your theme.

Steps for Plotting: Summary

A. Specify necessary layers

1. Data.
2. Aesthetics (*axis definitions*).

3. Geometric objects to use (*plot type*).

```
ggplot(data = ess10,
       aes(x = left_right)) +
  geom_histogram()
```

B. Customize your plot

1. Shape, color and size of geometric objects.
2. Fine-tune your x-scale and y-scale (axis ticks, ticks labels).
3. Label your axes, assign a title to your plot.
4. Choose a ggtheme.
5. Customize your theme.

ggplot2: Overview of Different Geoms

Histogram

- geom_histogram()

Densities

- geom_density(), stat_density()

Straight lines

- geom_abline(), geom_hline(), geom_vline()

Curves

- geom_curve()

Bar charts

- geom_bar(), geom_col(), stat_count()

Boxplots

- geom_boxplot(), stat_boxplot()

Dotplots

- geom_dotplot()

Scatterplots

- geom_point()

Violin plots

- geom_violin(), stat_ydensity()

Prerequisite: Data Wrangling Pipeline (I/III)

```r
library(tidyverse)
ess10 <- haven::read_dta("./dat/ESS10.dta")
ess10 <- ess10 %>% # subset variables
  select(country = cntry, # sociodemographics
         gender = gndr,
         education_years = eduyrs,
         trust_social = ppltrst, # multidimensional trust
         trust_parliament = trstprl,
         trust_legalSys = trstlgl,
         trust_police = trstplc,
         trust_politicians = trstplt,
         trust_parties = trstprt,
         trust_EP = trstep,
         trust_UN = trstun,
         left_right = lrscale, # attitudes
         life_satisfaction = stflife,
         pol_interest = polintr,
         voted = vote, # turnout
         party_choice = prtvtefr # party choice
         ) %>%
  mutate_at(c("country", "gender", "voted", "party_choice"), as.character) %>% # change types
  mutate_at("pol_interest", as.numeric) %>% # change types
  filter(country == "FR") # subset cases (only include France)
```

Prerequisite: Data Wrangling Pipeline (II/III)

```r
ess10 <- ess10 %>%
  mutate(gender = recode_factor(gender,
                                `1` = "Male",
                                `2` = "Female"),
       voted = recode_factor(voted,
                                `1` = "Yes",
                                `2` = "No",
                                `3` = "Not eligible"),
       party_choice = recode_factor(party_choice,
                                `1`  = "Lutte Ouvriére",
                                `2`  = "Nouv. Parti Anti-Capitaliste",
                                `3`  = "Parti Communiste Français",
                                `4`  = "La France Insoumise",
                                `5`  = "Parti Socialiste",
                                `6`  = "Europe Ecologie Les Verts",
                                `7`  = "La République en Marche",
                                `8`  = "Mouvement Démocrate",
                                `9`  = "Les Républicains",
                                `10` = "Debout la France",
                                `11` = "Front National",
                                `12` = "Other",
                                `13` = "Blank",
                                `14` = "Null")
  )
```

Prerequisite: Data Wrangling Pipeline (III/III)

```
ess10 <- ess10 %>%
  mutate(education_years = na_if(education_years, 114), # set 114 to missing
         pol_interest = (pol_interest * -1) + 5, # invert scale
         life_satisfaction = life_satisfaction + 1 # change scale to [1, 11]
         ) %>%
  drop_na(trust_politicians, gender, education_years,
          life_satisfaction, pol_interest) # list-wise deletion of missings
```

Constructing a Scatterplot

The European Social Survey Wave 10 asked people about their

- life satisfaction ranging from extremely dissatisfied to extremely satisfied
- trust in politicians ranging from no trust at all to complete trust

Measurement Life Satisfaction

> *All things considered, how satisfied are you with your life as a whole nowadays? Please answer using this card, where 0 means mextremely dissatisfied and 10 means extremely satisfied.*

Measurement Trust in Politicians

> *Using this card, please tell me on a score of 0-10 how much you personally trust each of the institutions I read out. 0 means you do not trust an institution at all, and 10 means you have complete trust.*

Constructing a Scatterplot

```
table(ess10$life_satisfaction)
```

```
##
##    1    2    3    4    5    6    7    8    9   10   11
##   35   15   43   53   81  191  187  351  502  227  222
```

```
table(ess10$trust_politicians)
```

```
##
##    0    1    2    3    4    5    6    7    8    9   10
##  188  100  220  280  255  428  240  126   51    8   11
```

```
cor(ess10$life_satisfaction, ess10$trust_politicians)
```
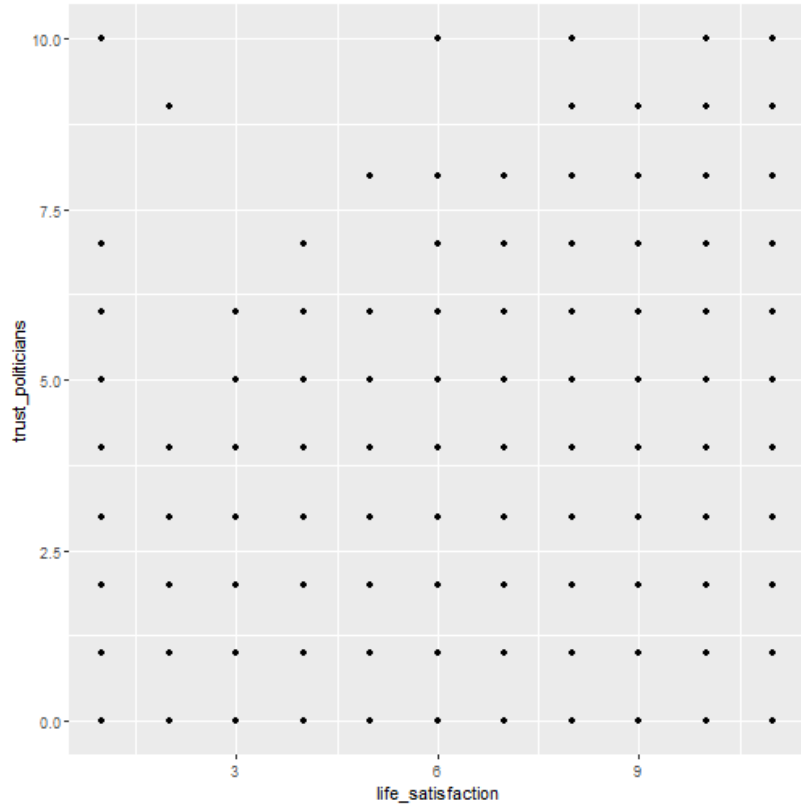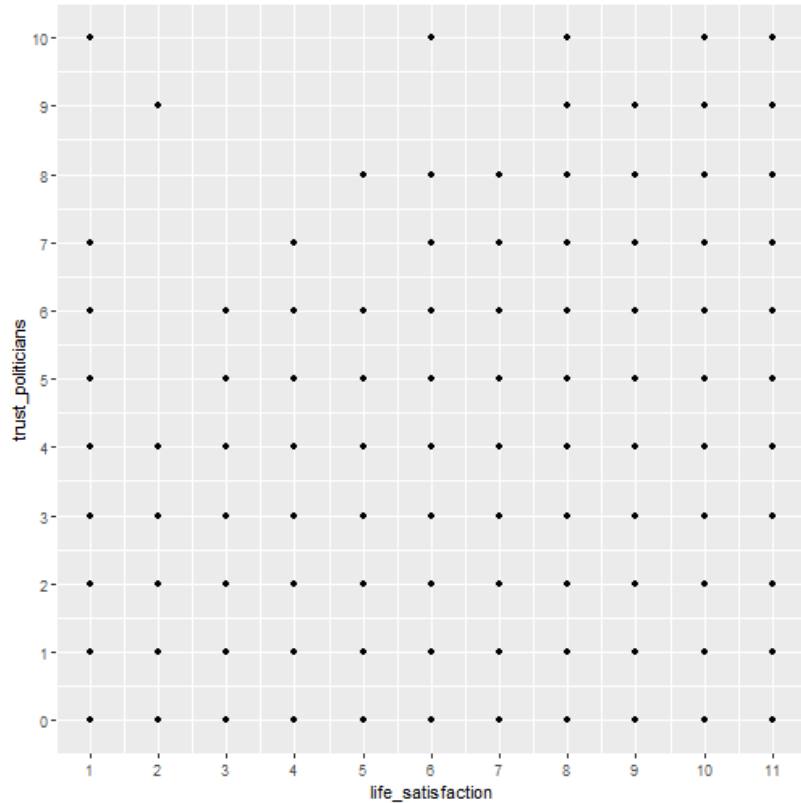
```
## [1] 0.26955
```

Constructing a Scatterplot
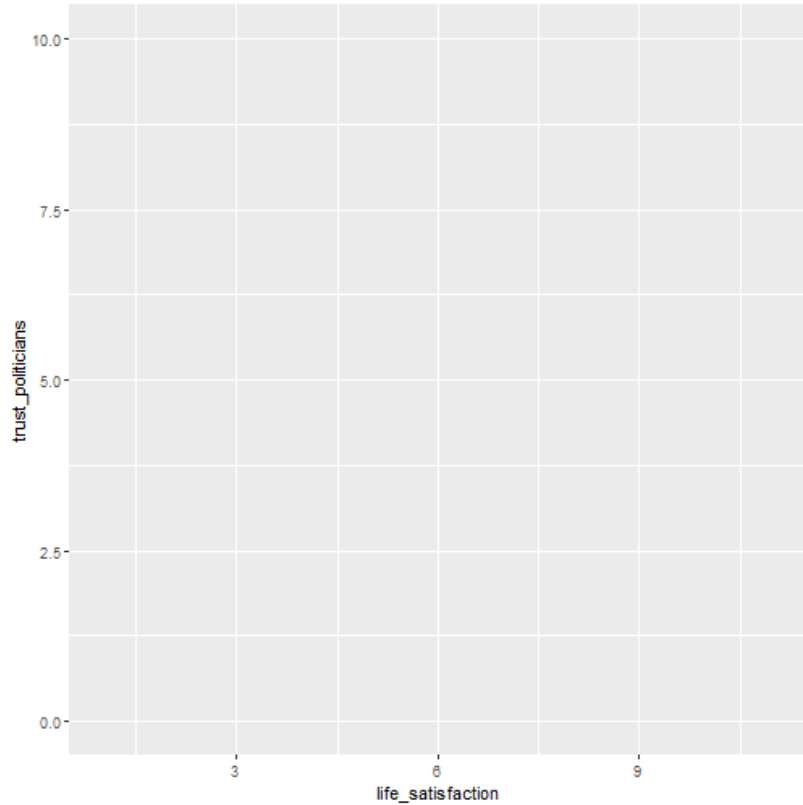

```
ggplot(data = ess10,
       aes(x = life_satisfaction,
           y = trust_politicians))
```

Constructing a Scatterplot


```
ggplot(data = ess10,
       aes(x = life_satisfaction,
           y = trust_politicians)) +
  geom_point()
```
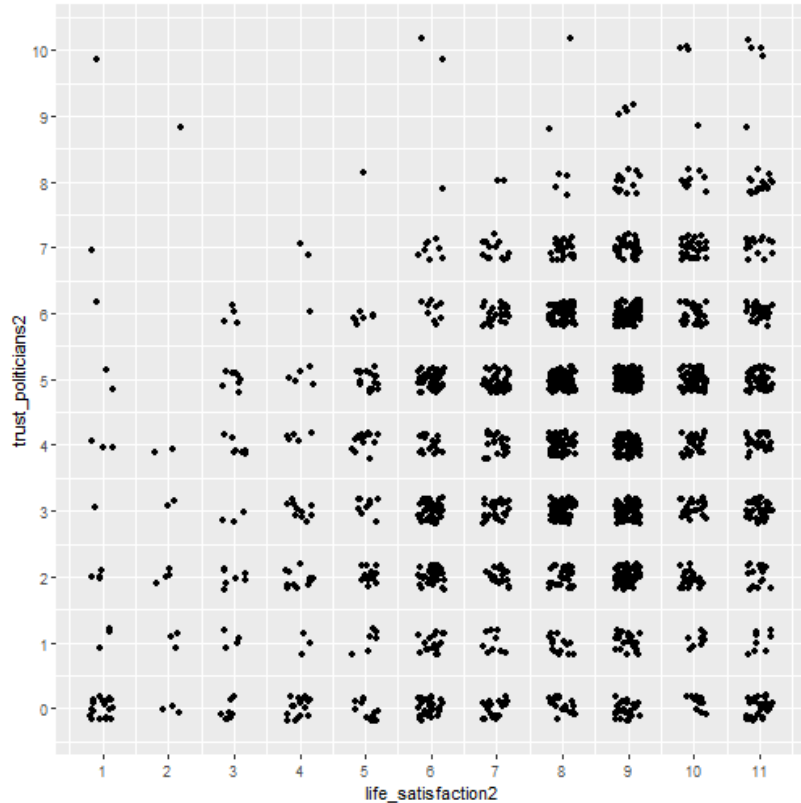
Constructing a Scatterplot


```
ggplot(data = ess10,
       aes(x = life_satisfaction,
           y = trust_politicians)) +
  geom_point() +
  scale_x_continuous(breaks = seq(1,11,1)) +
  scale_y_continuous(breaks = seq(0,10,1))
```
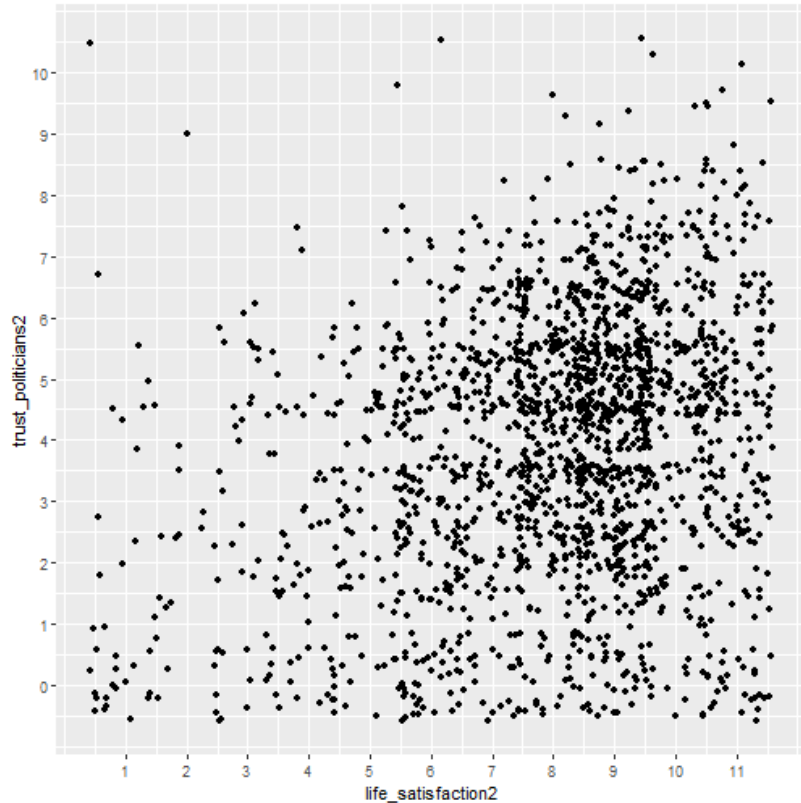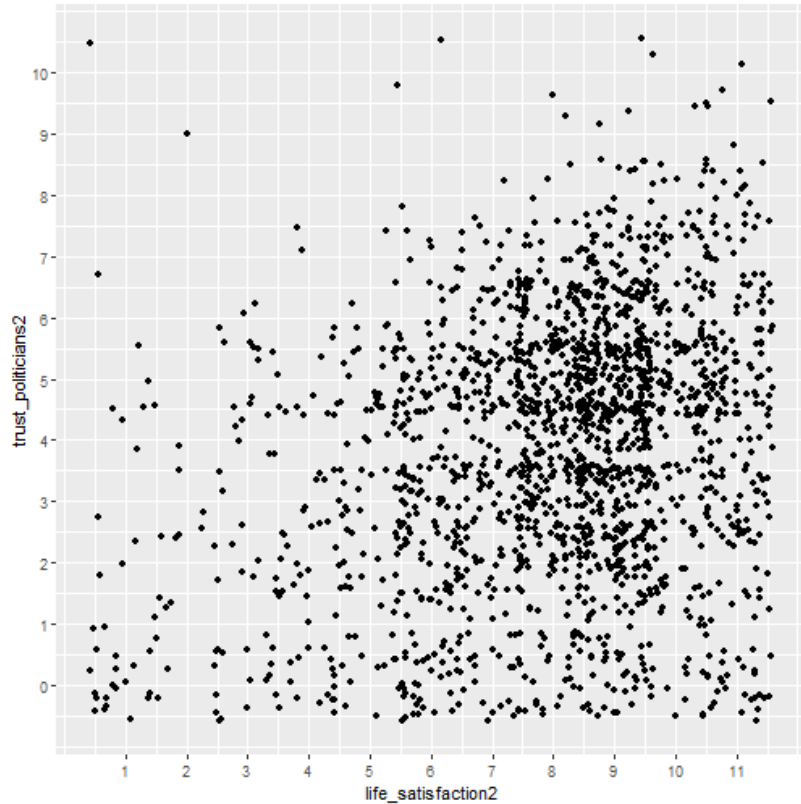
Constructing a Scatterplot


```
ess10$life_satisfaction2 <-
    jitter(ess10$life_satisfaction, 1)
ess10$trust_politicians2 <-
    jitter(ess10$trust_politicians, 1)

ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2)) +
  geom_point() +
  scale_x_continuous(breaks = seq(1,11,1)) +
  scale_y_continuous(breaks = seq(0,10,1))
```
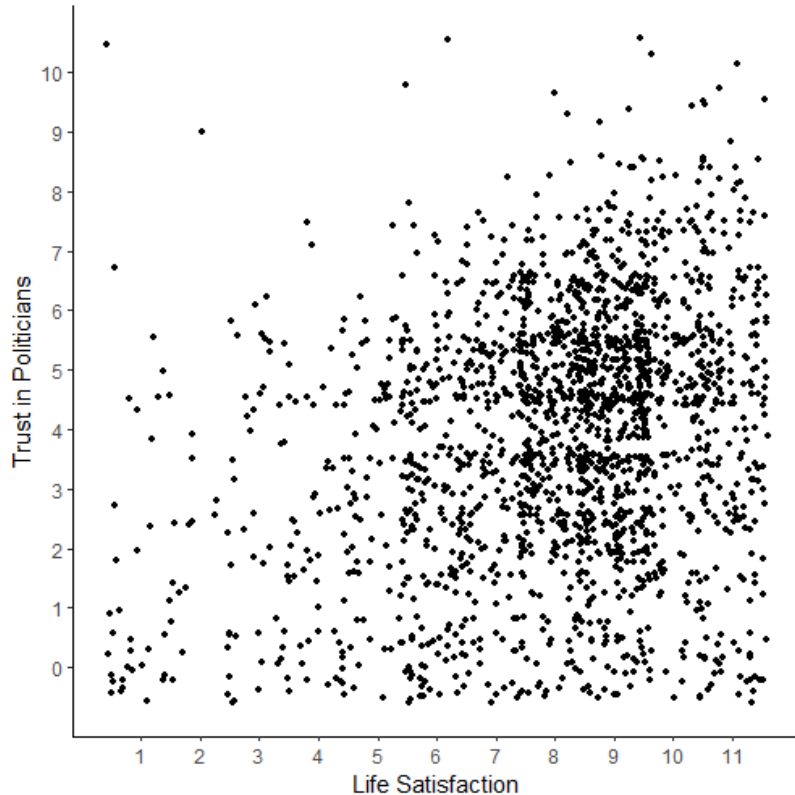
Constructing a Scatterplot


```
ess10$life_satisfaction2 <-
    jitter(ess10$life_satisfaction, 1)
ess10$trust_politicians2 <-
    jitter(ess10$trust_politicians, 1)

ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2)) +
  geom_point() +
  scale_x_continuous(breaks = seq(1,11,1)) +
  scale_y_continuous(breaks = seq(0,10,1))
```

Constructing a Scatterplot


```
ess10$life_satisfaction2 <-
    jitter(ess10$life_satisfaction, 3)
ess10$trust_politicians2 <-
    jitter(ess10$trust_politicians, 3)

ggplot(data = ess10,
        aes(x = life_satisfaction2,
            y = trust_politicians2)) +
    geom_point() +
    scale_x_continuous(breaks = seq(1,11,1)) +
    scale_y_continuous(breaks = seq(0,10,1))
```
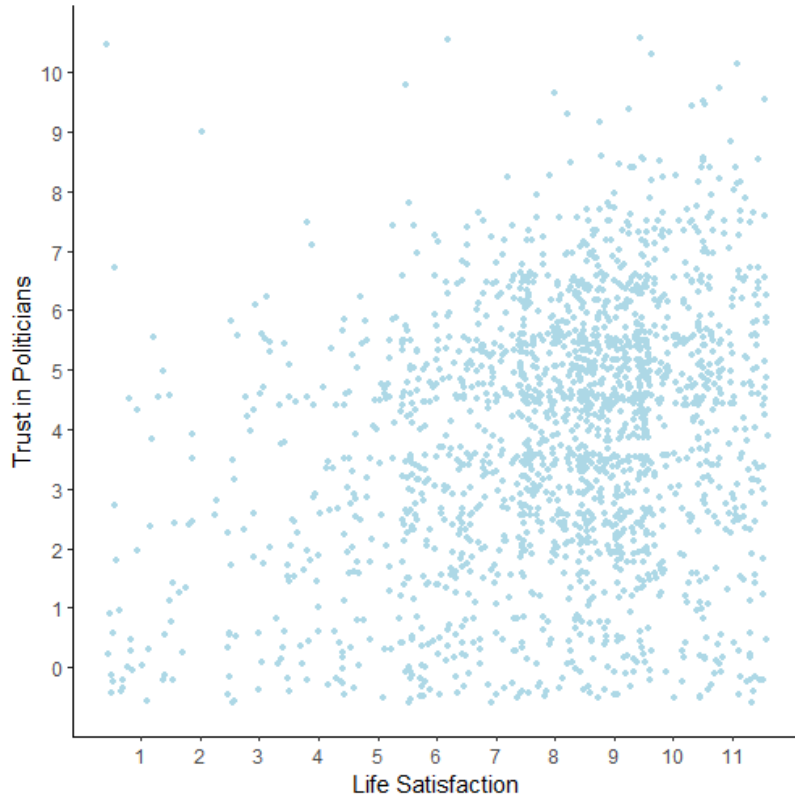
Constructing a Scatterplot
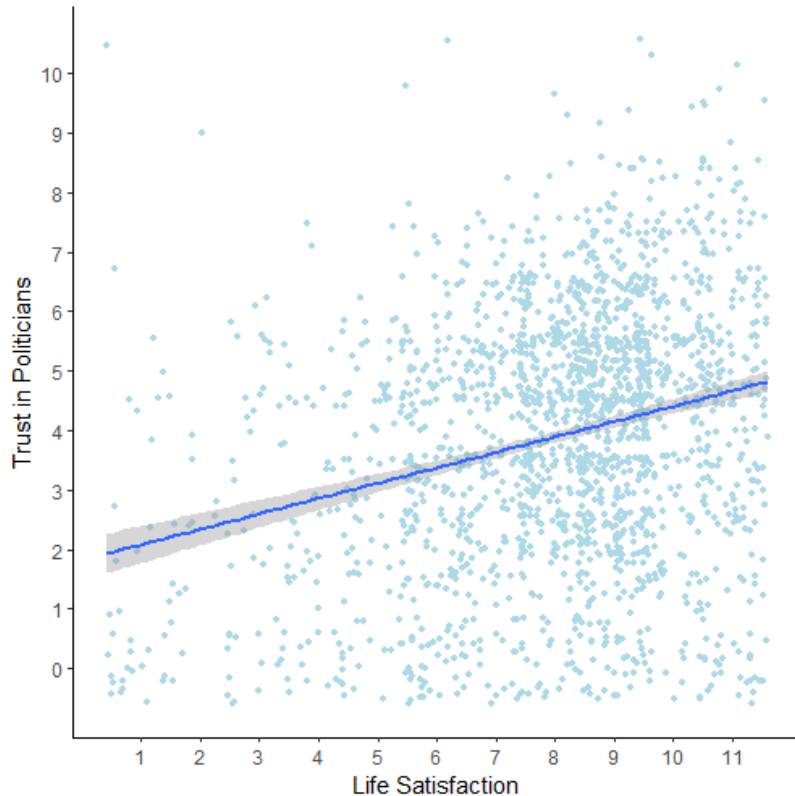


```
ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2)) +
  geom_point() +
  scale_x_continuous(breaks = seq(1,11,1)) +
  scale_y_continuous(breaks = seq(0,10,1))
```

Constructing a Scatterplot


```
ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2)) +
  geom_point() +
  scale_x_continuous(breaks = seq(1,11,1)) +
  scale_y_continuous(breaks = seq(0,10,1)) +
  labs(x = "Life Satisfaction",
       y = "Trust in Politicians") +
  theme_classic(base_size = 15)
```

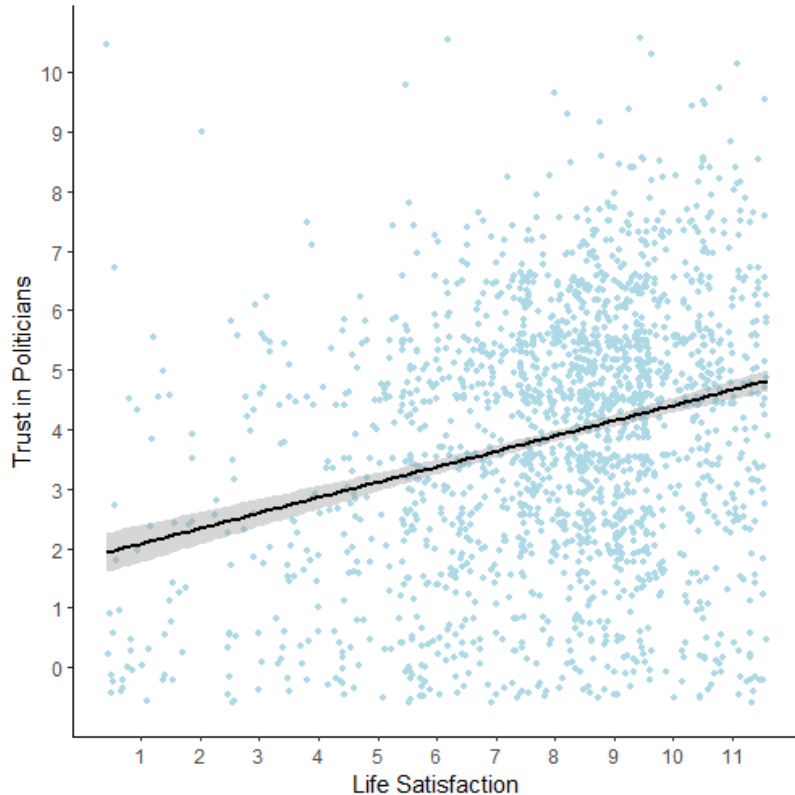# Constructing a Scatterplot


```
ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2)) +
  geom_point(col = "lightblue") +
  scale_x_continuous(breaks = seq(1,11,1)) +
  scale_y_continuous(breaks = seq(0,10,1)) +
  labs(x = "Life Satisfaction",
       y = "Trust in Politicians") +
  theme_classic(base_size = 15)
```

Constructing a Scatterplot


```
ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2)) +
  geom_point(col = "lightblue") +
  geom_smooth(method = "lm") +
  scale_x_continuous(breaks = seq(1,11,1)) +
  scale_y_continuous(breaks = seq(0,10,1)) +
  labs(x = "Life Satisfaction",
       y = "Trust in Politicians") +
  theme_classic(base_size = 15)
```

Constructing a Scatterplot


```
ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2)) +
  geom_point(col = "lightblue") +
  geom_smooth(method = "lm", col = "black") +
  scale_x_continuous(breaks = seq(1,11,1)) +
  scale_y_continuous(breaks = seq(0,10,1)) +
  labs(x = "Life Satisfaction",
       y = "Trust in Politicians") +
  theme_classic(base_size = 15)
```

Constructing a Scatterplot

What if the relationship between life satisfaction and trust in politicians is moderated by political interest?
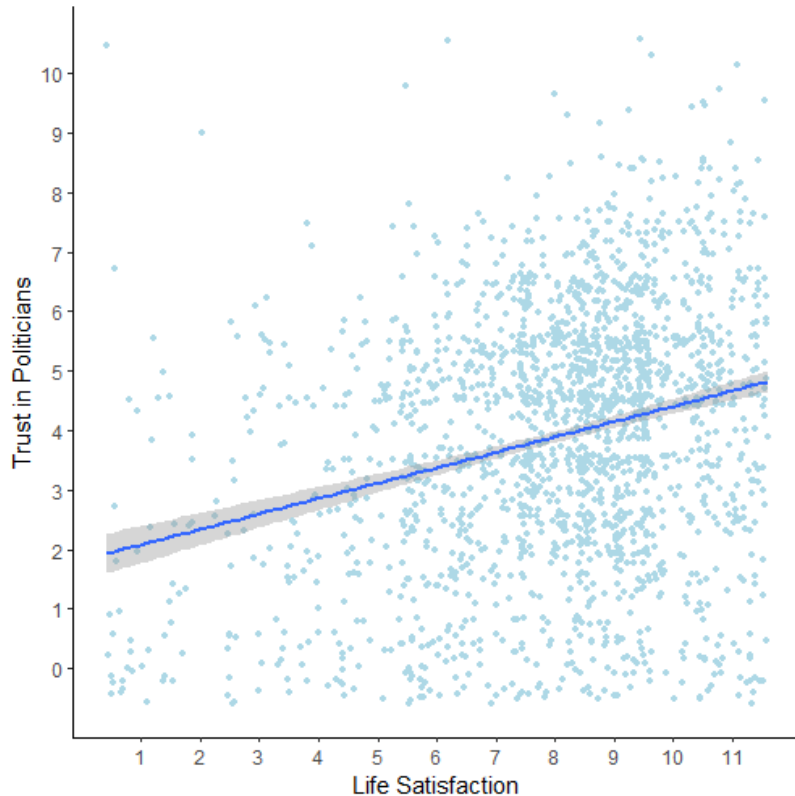
- Those who are politically interested might trust politicians either way
- For those that are not politically interested, life satisfaction might matter more

```
ess10 <- ess10 %>%
  mutate(pol_interestBIN = recode_factor(
    pol_interest,
    `1` = 0,
    `2` = 0,
    `3` = 1,
    `4` = 1)
  )
```

```
table(ess10$pol_interest, ess10$pol_interestBIN)
```
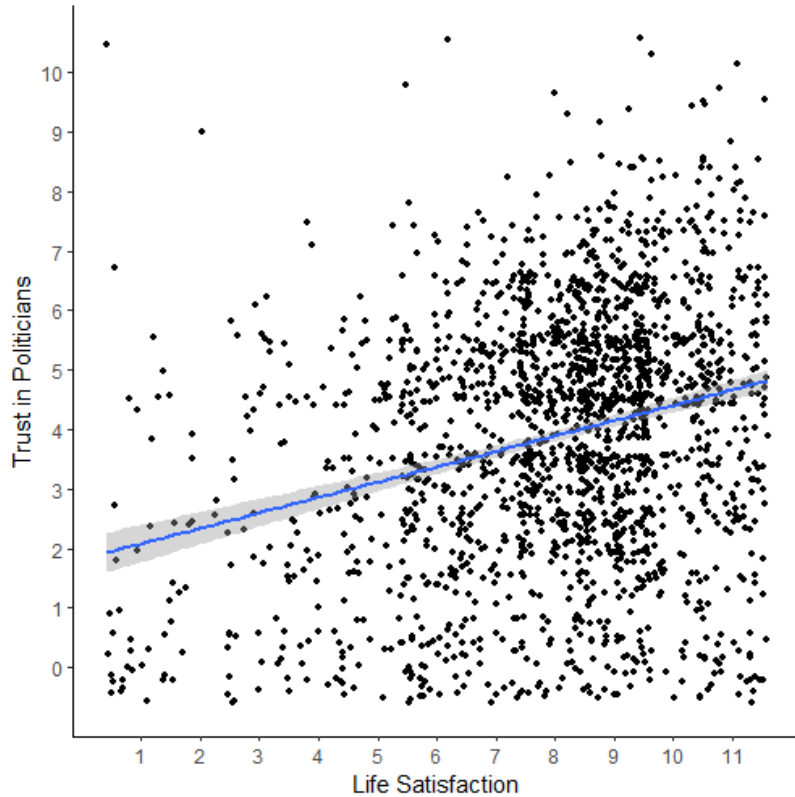
```
##
##       0   1
##   1 373   0
##   2 770   0
##   3   0 469
##   4   0 295
```

Constructing a Scatterplot
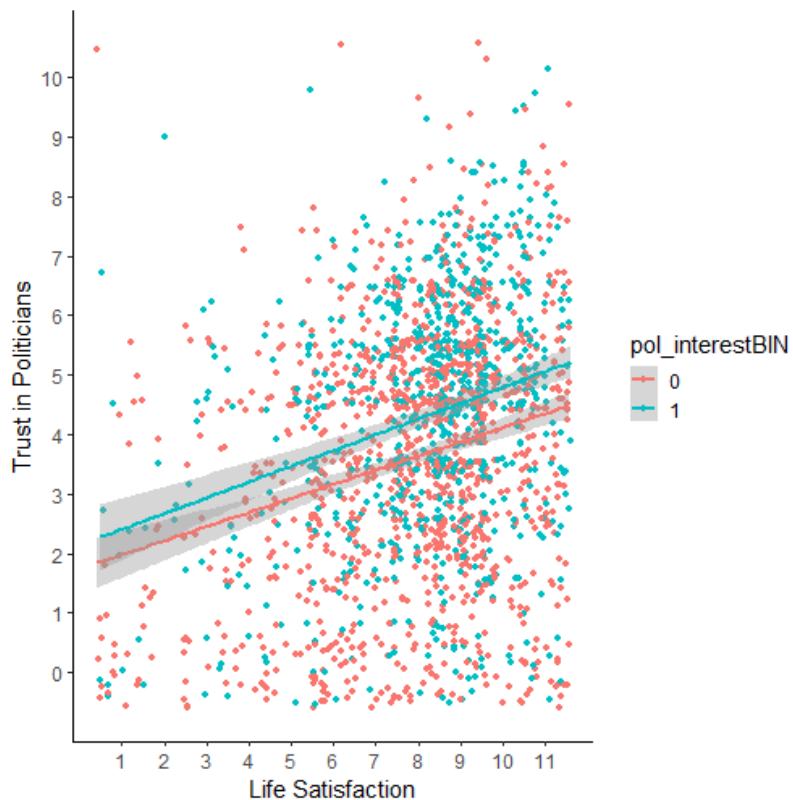


```
ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2)) +
  geom_point(col = "lightblue") +
  geom_smooth(method = "lm") +
  scale_x_continuous(breaks = seq(1,11,1)) +
  scale_y_continuous(breaks = seq(0,10,1)) +
  labs(x = "Life Satisfaction",
       y = "Trust in Politicians") +
  theme_classic(base_size = 15)
```
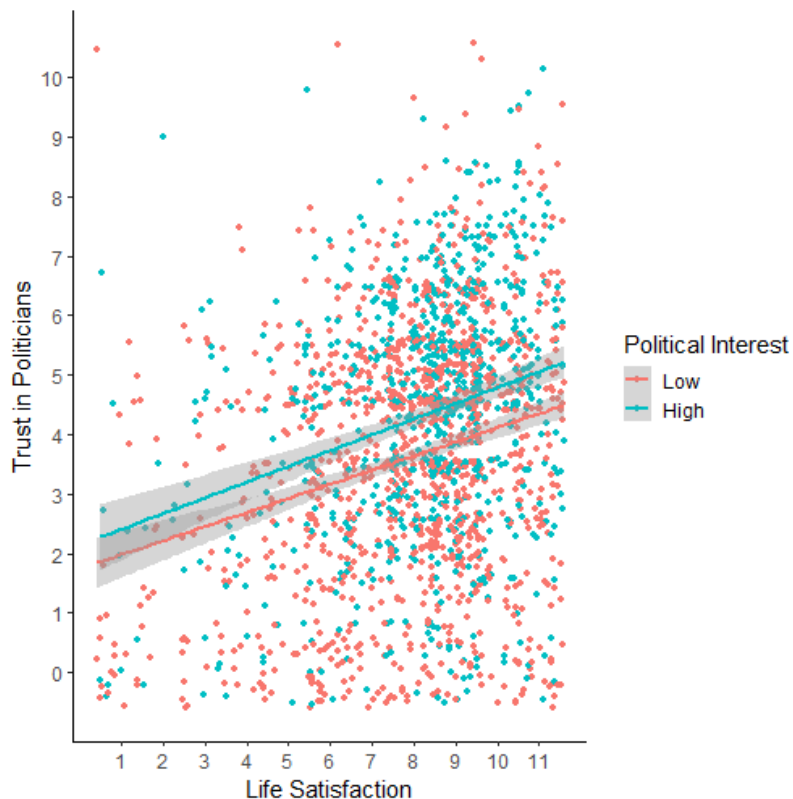
Constructing a Scatterplot


```
ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2)) +
geom_point() +
geom_smooth(method = "lm") +
scale_x_continuous(breaks = seq(1,11,1)) +
scale_y_continuous(breaks = seq(0,10,1)) +
labs(x = "Life Satisfaction",
     y = "Trust in Politicians") +
theme_classic(base_size = 15)
```

Constructing a Scatterplot


```r
ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2,
           color = pol_interestBIN)) +
geom_point() +
geom_smooth(method = "lm") +
scale_x_continuous(breaks = seq(1,11,1)) +
scale_y_continuous(breaks = seq(0,10,1)) +
labs(x = "Life Satisfaction",
     y = "Trust in Politicians") +
theme_classic(base_size = 15)
```

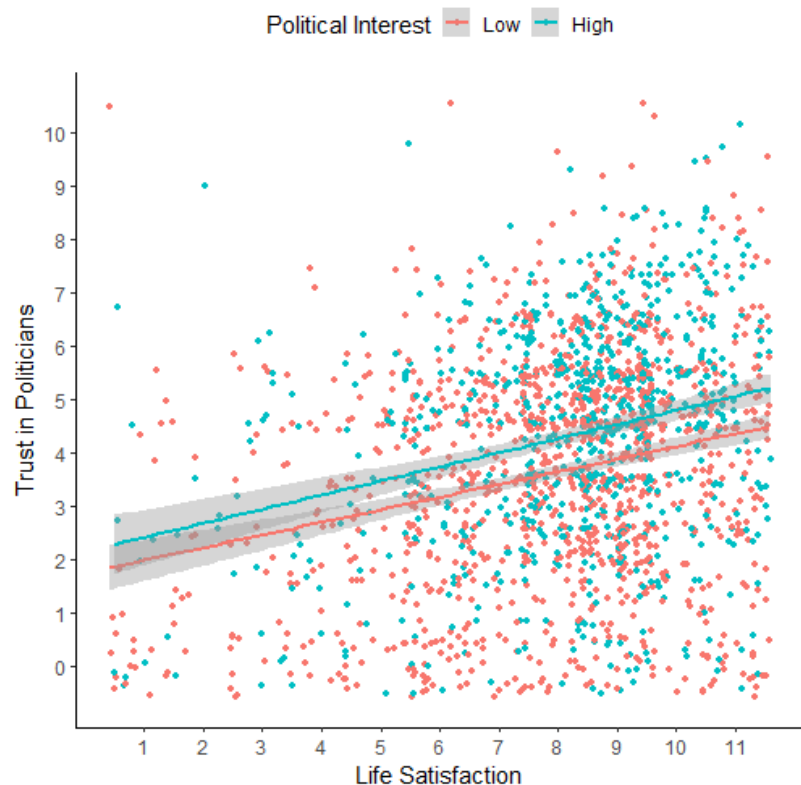# Constructing a Scatterplot


```
ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2,
           color = pol_interestBIN)) +
geom_point() +
geom_smooth(method = "lm") +
scale_x_continuous(breaks = seq(1,11,1)) +
scale_y_continuous(breaks = seq(0,10,1)) +
labs(x = "Life Satisfaction",
     y = "Trust in Politicians",
     color="Political Interest") +
theme_classic(base_size = 15) +
scale_color_discrete(
    labels = c("Low", "High"))
```

Constructing a Scatterplot
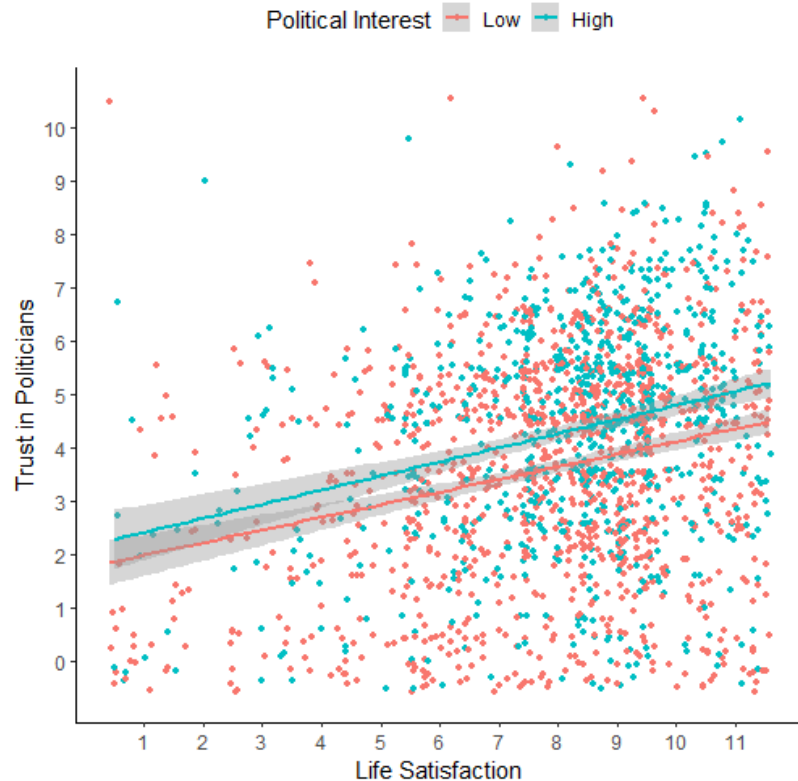


```
ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2,
           color = pol_interestBIN)) +
  geom_point() +
  geom_smooth(method = "lm") +
  scale_x_continuous(breaks = seq(1,11,1)) +
  scale_y_continuous(breaks = seq(0,10,1)) +
  labs(x = "Life Satisfaction",
       y = "Trust in Politicians",
       color="Political Interest") +
  theme_classic(base_size = 15) +
  scale_color_discrete(
    labels = c("Low", "High")) +
  theme(legend.position = "top")
```

Constructing a Scatterplot


```r
ggplot(data = ess10,
       aes(x = life_satisfaction2,
           y = trust_politicians2,
           color = pol_interestBIN)) +
  geom_point() +
  geom_smooth(method = "lm") +
  scale_x_continuous(breaks = seq(1,11,1)) +
  scale_y_continuous(breaks = seq(0,10,1)) +
  labs(x = "Life Satisfaction",
       y = "Trust in Politicians",
       color="Political Interest") +
  theme_classic(base_size = 15) +
  scale_color_discrete(
    labels = c("Low", "High")) +
  theme(legend.position = "top")
```

References

Parts of this course are inspired by the following resources:

- Wickham, Hadley and Garrett Grolemund, 2017. *R for Data Science - Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly.

- Bahnsen, Oke and Guido Ropers, 2022. *Introduction to R for Quantitative Social Science*. Course held as part of the GESIS Workshop Series.

- Breuer, Johannes and Stefan Jünger, 2021. *Introduction to R for Data Analysis*. Course held as part of the GESIS Summer School in Survey Methodology.

- Teaching material developed by Verena Kunz, David Weyrauch, Oliver Rittmann and Viktoriia Semenova.