PSC4375: Boxplots and QQ-plots

Week 4: Lecture 9

Prof. Weldzius

Villanova University

Slides Updated: 2025-01-26

Assassination attempts

- Load the assassination attempts data; see the possible attempt results

```
library(tidyverse)
data(leaders, package = "qss")
unique(leaders$result)
```

```
##  [1] "not wounded"
##  [2] "dies within a day after the attack"
##  [3] "survives, whether wounded unknown"
##  [4] "wounded lightly"
##  [5] "plot stopped"
##  [6] "hospitalization but no permanent disability"
##  [7] "dies between a day and a week"
##  [8] "dies, timing unknown"
##  [9] "survives but wounded severely"
## [10] "dies between a week and a month"
```

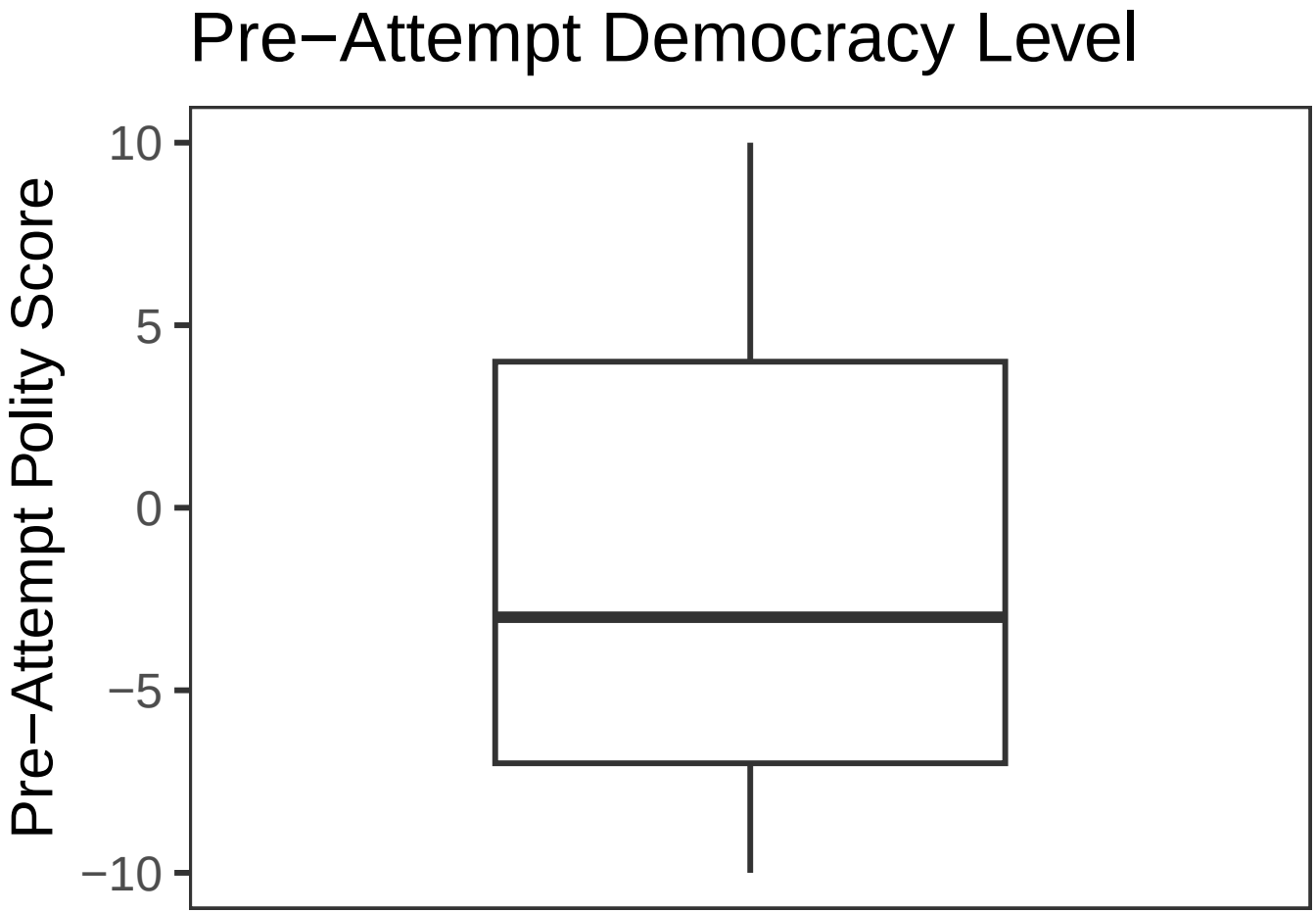
Creating an attempt fatal variable

- use `ifelse` to create a `fatal` variable

```
## create new vector of unique results of "result"
lev <- unique(leaders$result)
leaders <- leaders %>%
  mutate(fatal = ifelse(result %in% lev[c(2,7,8,10)], 1,0))
leaders %>%
  summarize(mean(fatal))
```

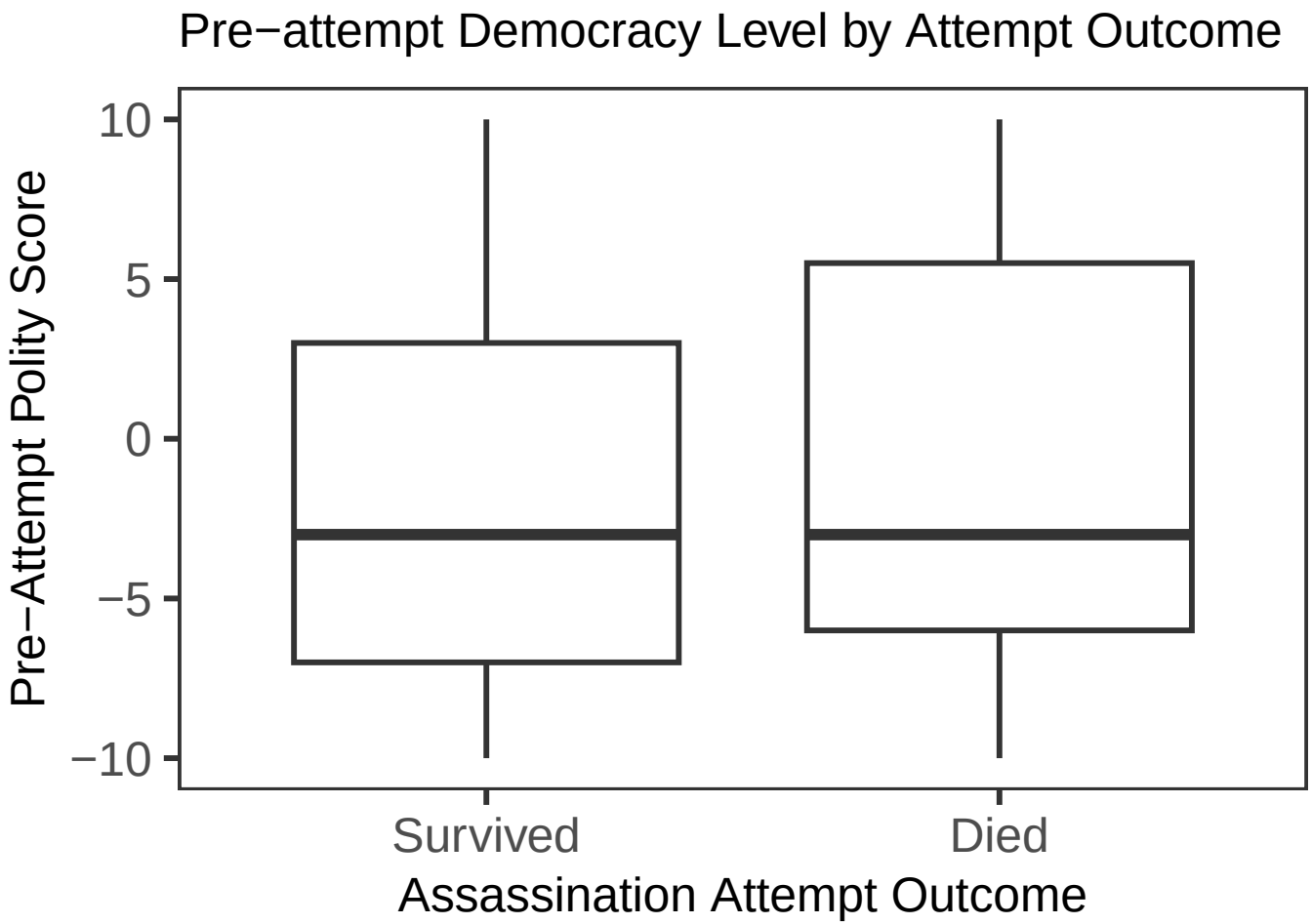
```
##   mean(fatal)
## 1       0.216
```

Remember boxplots?



Comparing distribution with the boxpot

- What if we want to know how the distribution varies by success?



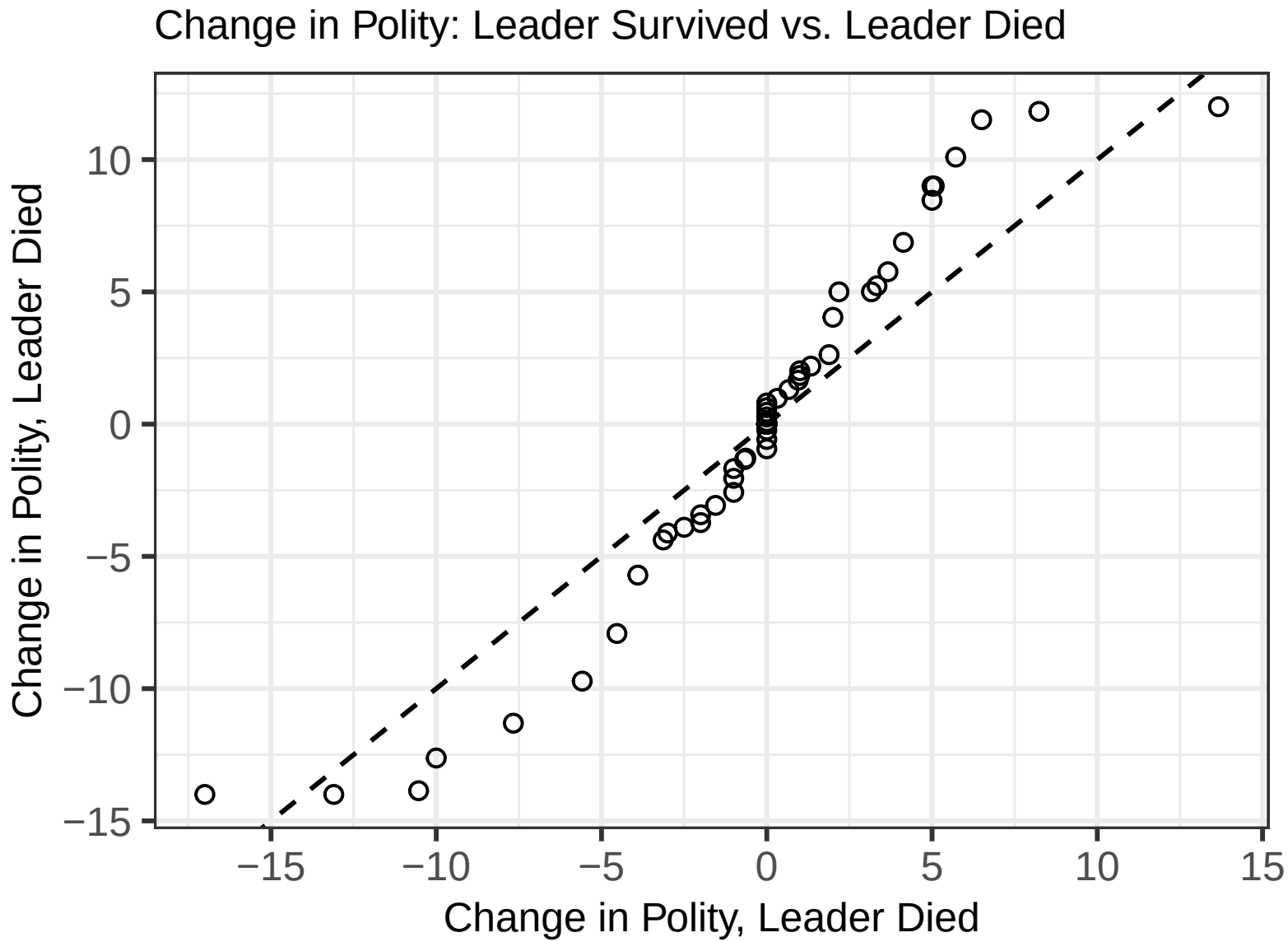
Boxplot comparisons in R

```
leaders %>%
  ggplot(aes(y = politybefore,
             x = factor(fatal,labels = c("Survived", "Died")))
  geom_boxplot() +
  scale_y_continuous(breaks = seq(-10, 10,by = 5)) +
  labs(title = "Pre-attempt Democracy Level by Attempt Outcome
       y = "Pre-Attempt Polity Score",
       x = "Assassination Attempt Outcome") +
  theme_bw() +
  theme(plot.title = element_text(size=9),
        axis.title.x = element_text(size = 9),
        axis.title.y = element_text(size = 9),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())
```

Quantile-Quantile Plot

- How do we compare distributions of two variables that are not in the same dataset?
  - Could use boxplots, but it's only a crude summary of the distributions.
- **Quantile-quantile plot (Q-Q plot)**: scatterplot of **quantiles**
  - (min of X, min of Y)
  - (median of X, median of Y)
  - (25th percentile of X, 25th percentile of Y)
- Intuitions:
  - If distributions are the same $\rightsquigarrow$ all points on a 45-degree line
  - Points above 45° line $\rightsquigarrow$ y-axis variable has larger value of the quantile
  - Point below 45° line $\rightsquigarrow$ x-axis variable has larger value of the quantile
  - Steeper slope than 45° line $\rightsquigarrow$ y-axis variable has more spread
  - Flatter slope than 45° line $\rightsquigarrow$ x-axis variable has more spread

QQ-plot example



QQ-plot example (setup)

```
## calculate change in polity
leaders <- leaders %>%
  mutate(polity_change = polityafter - politybefore)

## set quantile vectors
quantile_probs <- seq(from = 0, to = 1, by = 0.01)
quantile_names <- as.character(quantile_probs)

## generate dataframe for plot
quantiles <- leaders %>%
  group_by(fatal) %>%
  summarize(politychng_quantile = quantile(polity_change, prob
            quantile = quantile_names) %>%
  pivot_wider(names_from = fatal,
              values_from = politychng_quantile)
```

QQ-plot example (plot)

```
quantiles %>%
  ggplot(aes(x = `0`, y = `1`)) +
  geom_point(shape = 1) +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed") +
  scale_y_continuous(breaks = seq(-20, 15,by = 5)) +
  scale_x_continuous(breaks = seq(-20, 15,by = 5)) +
  labs(title = "Change in Polity: Leader Survived vs. Leader D
       y = "Change in Polity, Leader Died",
       x = "Change in Polity, Leader Died") +
  theme_bw() +
  theme(plot.title = element_text(size=9),
        axis.title.x = element_text(size = 9),
        axis.title.y = element_text(size = 9))
```