

# Playing with Data II: Data Visualization



Rebeca J. Agosto Rosa  
[rja2@illinois.edu](mailto:rja2@illinois.edu)

# Objectives

- Why is data visualization important?
- What makes plots effective?
- How do we make plots in R?

Why is  
data visualization  
important?



# Data analysis

Good data analysis involves both **numeric** and **visual** representations of the data at different stages:

1. Data exploration
2. Model building
3. Assessment of model fit
4. Reporting of the results

Quick review:

Numerical representations of data

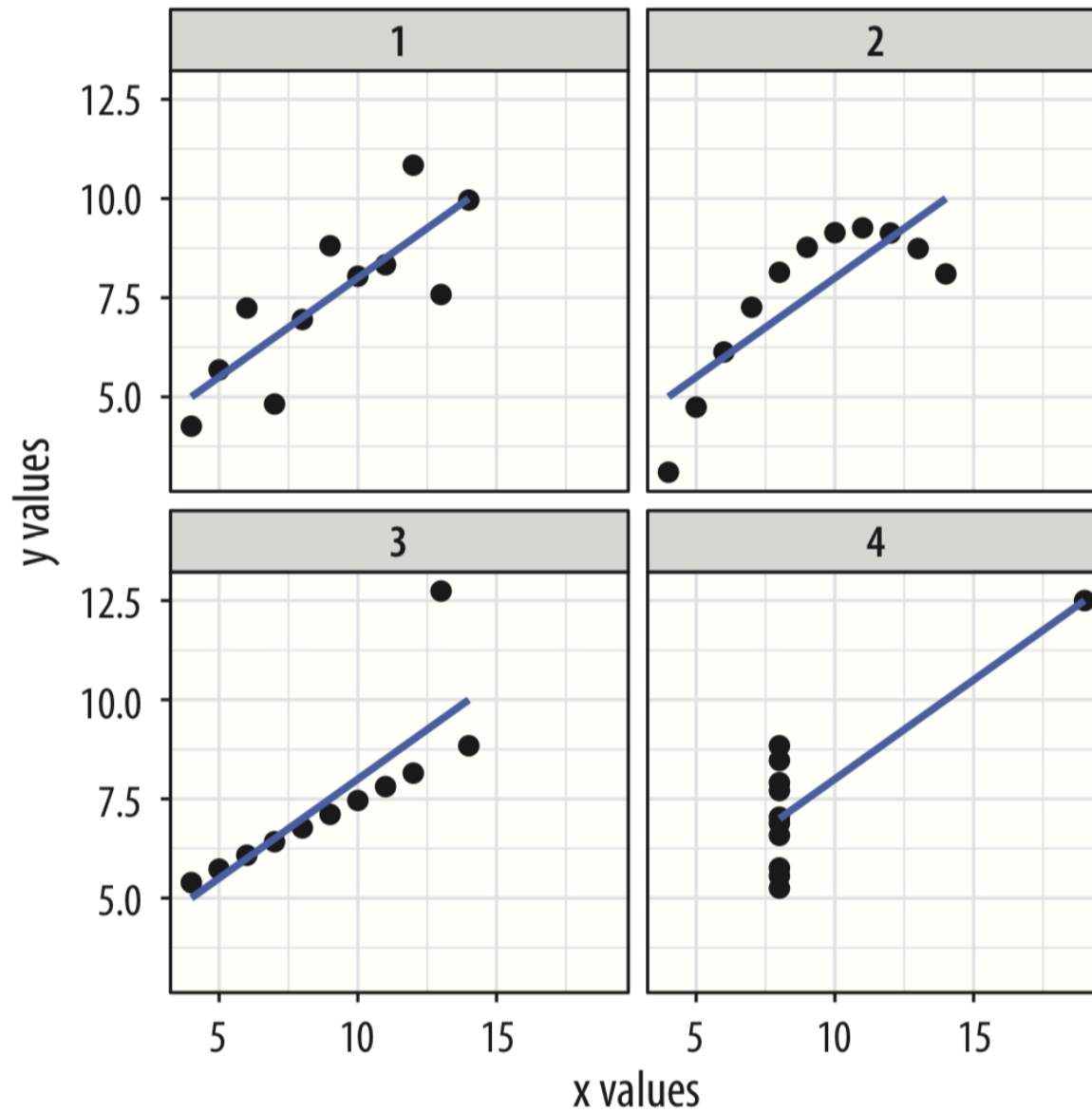
Deliberate simplification of data

- Summary statistics  
e.g., central tendency measures
- Correlations
- Model estimates  
e.g., coefficients in the regression table

# Visual representations of data

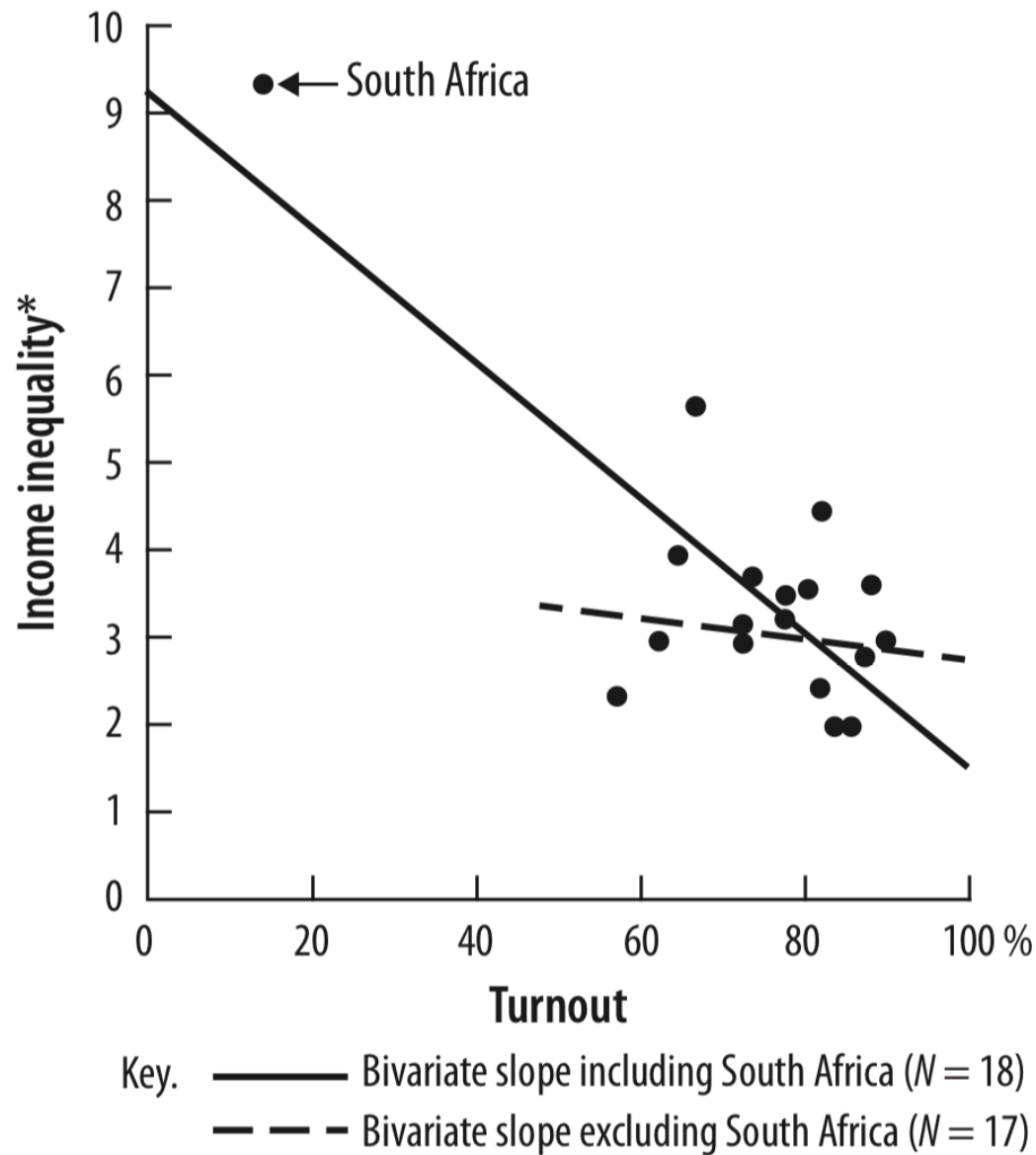
## Some benefits

- Observe trends
- Uncover unexpected distributions
- Detect potential outliers
- Identify potential measurement errors



**Figure 1.1:** Plots of Anscombe's quartet.

Source:  
Healy (2018), ch. 1



**Figure 1.2:** Seeing the effect of an outlier on a regression line.

Source:  
Healy (2018), ch. 1

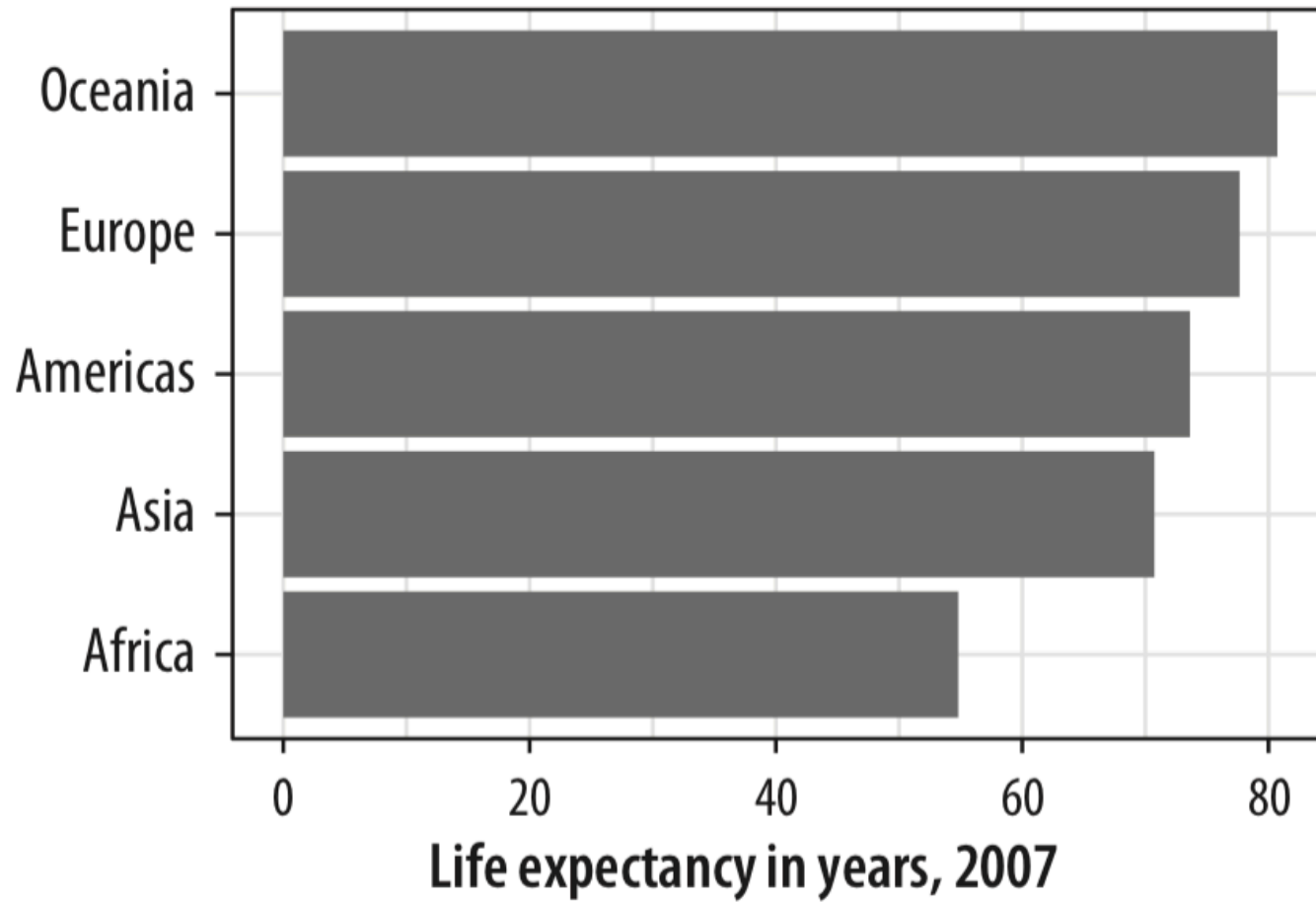


# Making effective plots





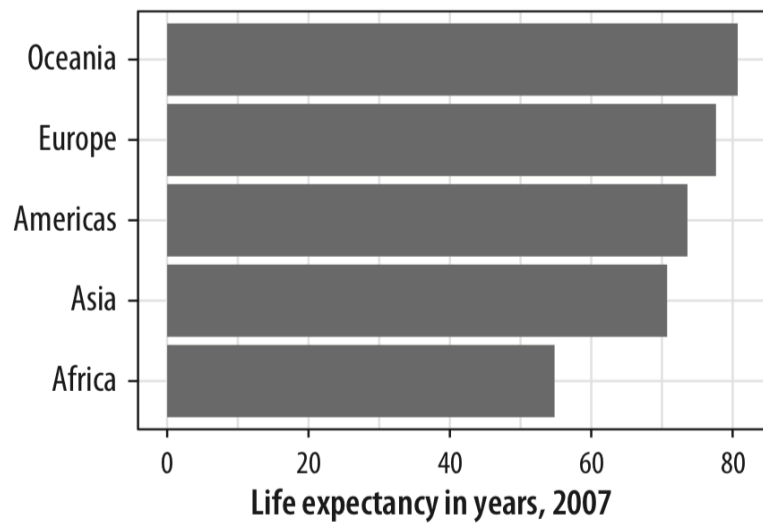
Source:  
Healy (2018), ch. 1



Source:  
Healy (2018), ch. 1



Source:  
Healy (2018), ch. 1



Source:  
Healy (2018), ch. 1

# Key Considerations (Healy 2018)

- **Aesthetic**  
Elegant design choices
- **Substantive**  
Accurate & clear data representation
- **Perceptual**  
How the brain processes visual information matters

**We need to be honest!**

# Making Plots in R



# The Basics

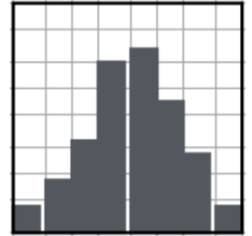
There are **many types** of graphs, and **many ways** to make them.

Our focus:

1. Plotting one and two variables
2. R base functions
3. ggplot2 package

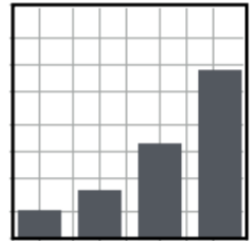


# Graphs to explore data



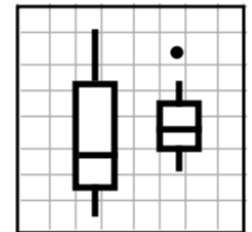
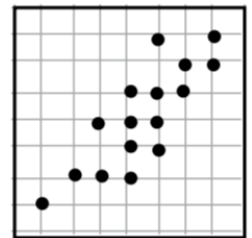
## One variable

- Histogram – continuous (numeric)
- Bar plot – discrete (categorical)



## Two variables

- Scatter plot – two continuous
- Boxplot – one continuous, one discrete



# R base functions

Histogram

```
hist(data$variable)
```

Bar plot

```
barplot(table(data$variable))
```

Scatter plot

```
plot(x, y, data)
```

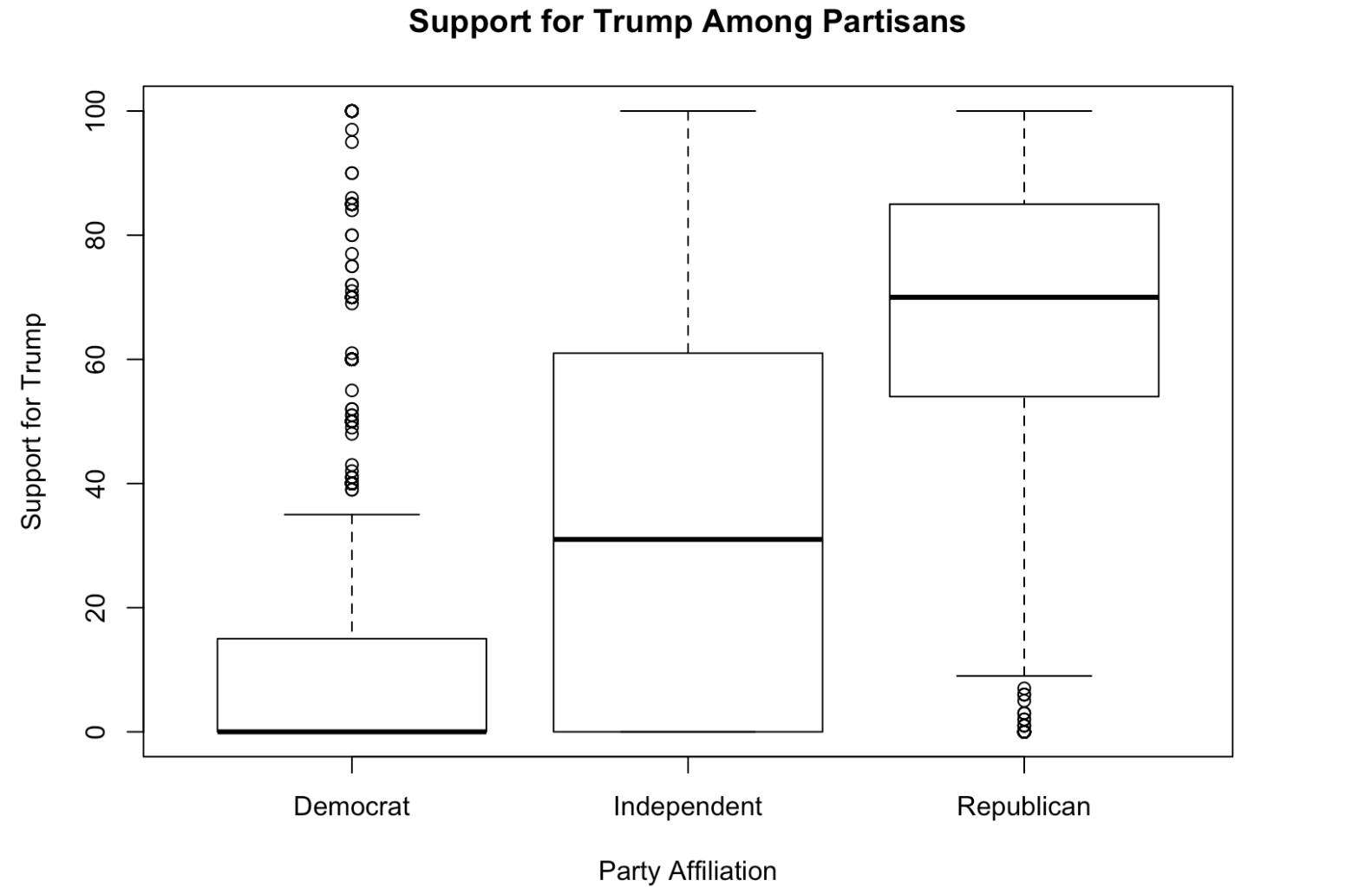
Boxplot

```
boxplot(y ~ group, data)
```

y = continuous variable

group = discrete variable

```
boxplot(feelTrump ~ partyID, data = anes16,
        main = "Support for Trump Among Partisans",
        xlab = "Party Affiliation",
        ylab = "Support for Trump")
```



# ggplot2

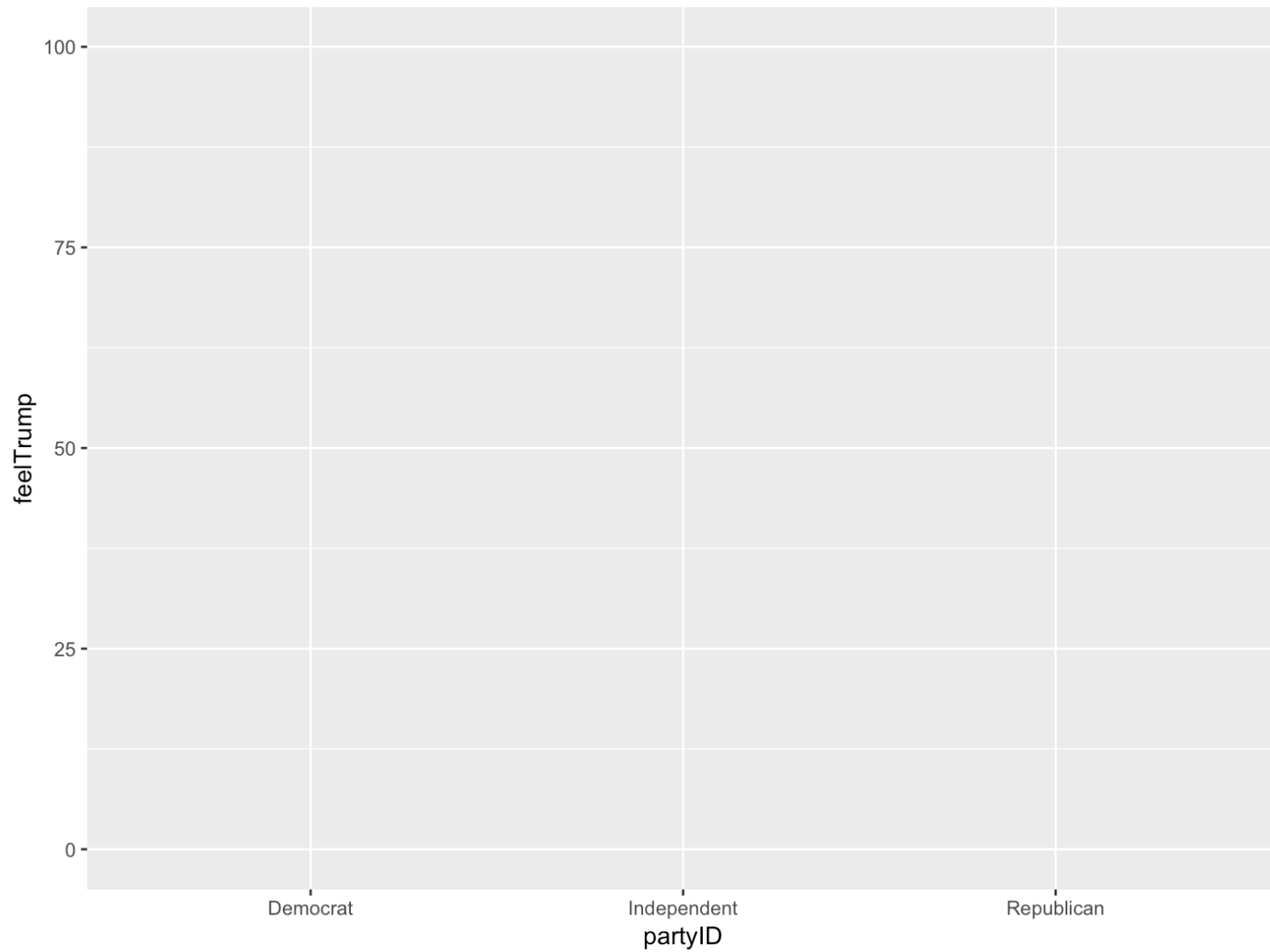
R package; system to create graphics

Based on *The Grammar of Graphics*

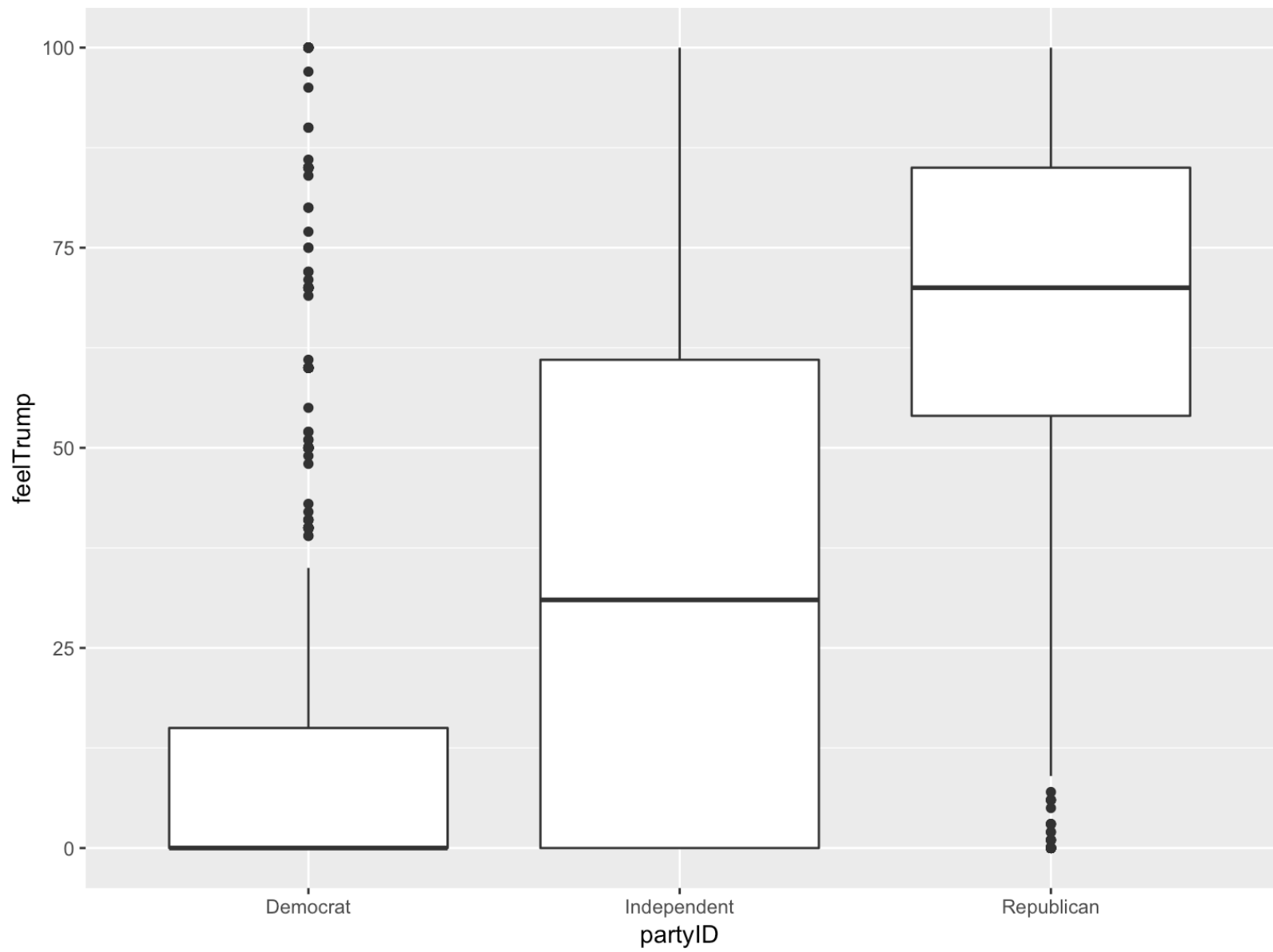
Graphs are built layer by layer

1. Specify the base:
  - a. **data** to be used
  - b. **mapping rules** (how the data will be displayed in a **coordinate system**)
2. Specify layers:
  - a. “**geoms**” (geometric objects): how to represent the data (e.g., points, lines, bars)
  - b. scales, titles, themes

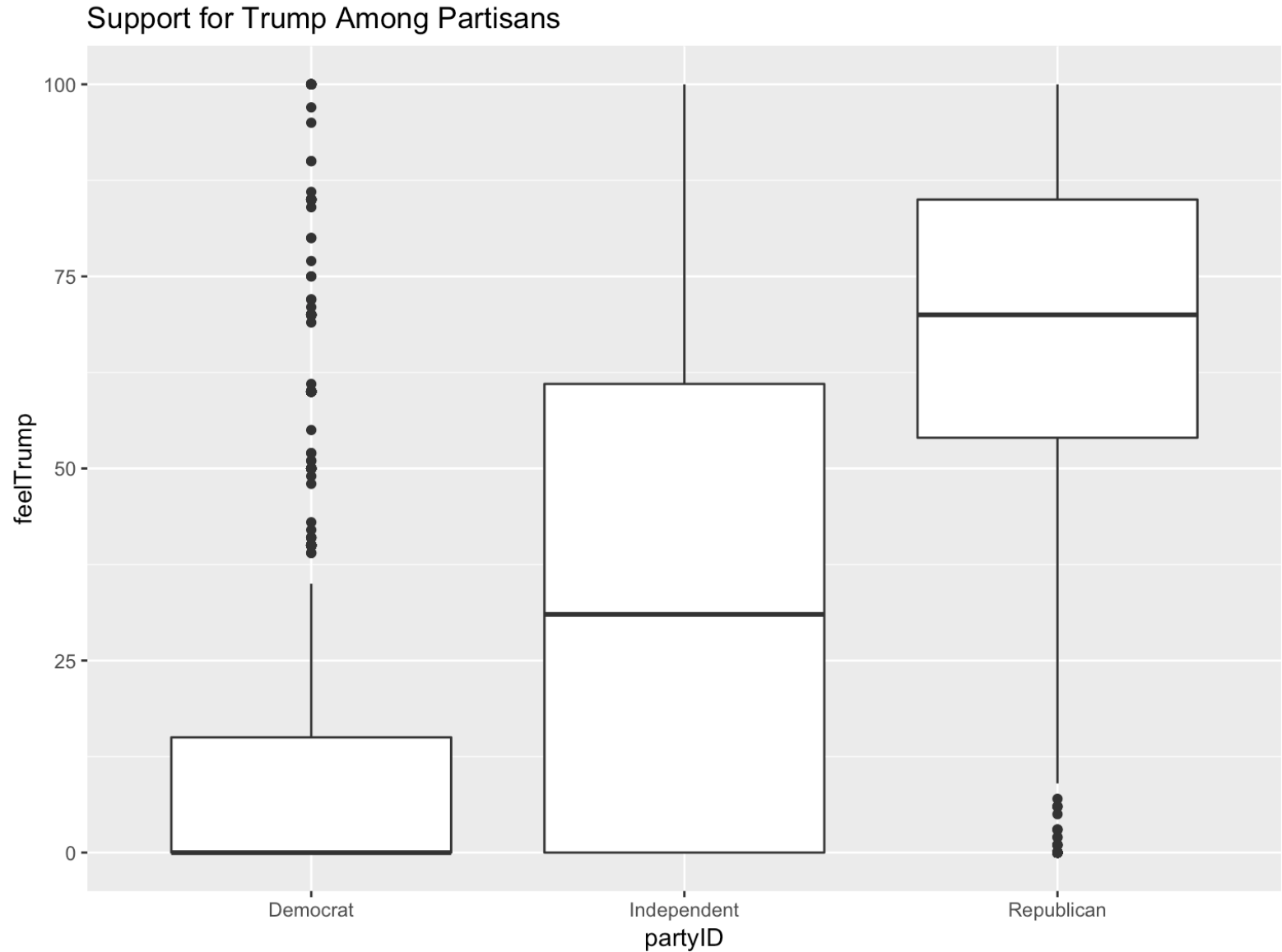
```
library(ggplot2)
p <- ggplot(data = anes16, aes(x = partyID, y = feelTrump))
p
```



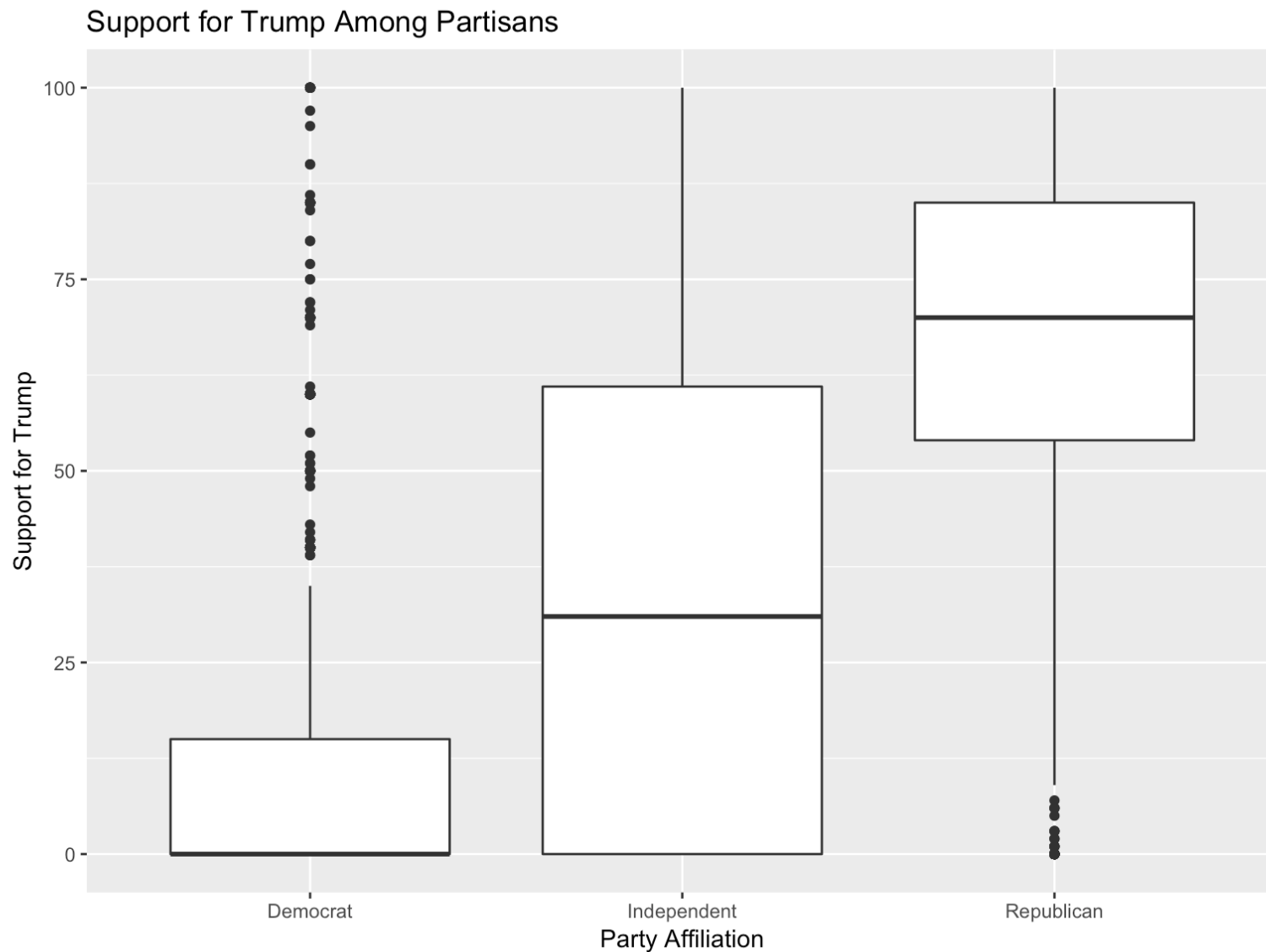
```
p + geom_boxplot()
```



```
p + geom_boxplot() +  
  labs(title = "Support for Trump Among Partisans")
```

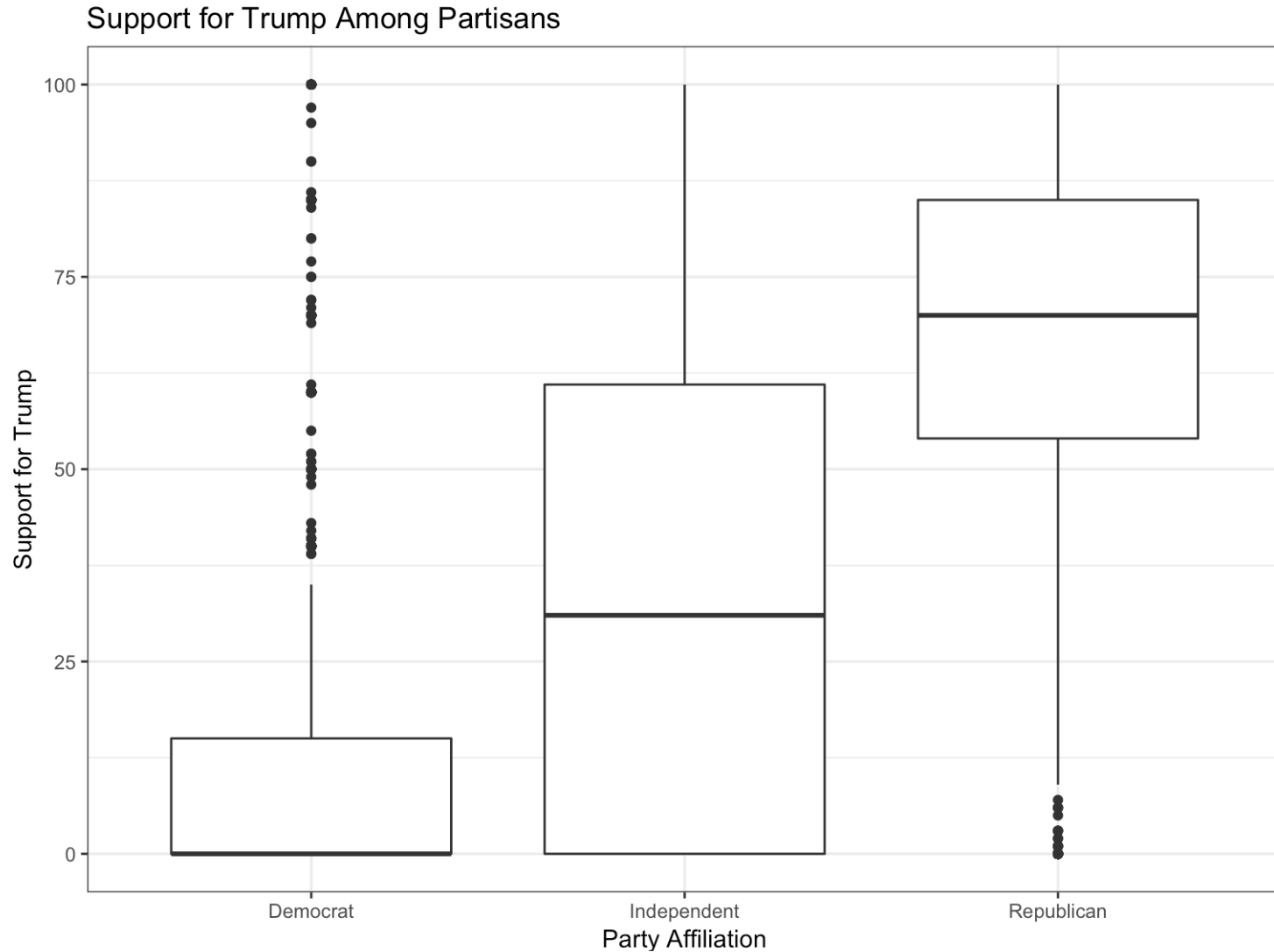


```
p + geom_boxplot() +  
  labs(title = "Support for Trump Among Partisans",  
       x = "Party Affiliation", y = "Support for Trump")
```





```
p + geom_boxplot() +  
  labs(title = "Support for Trump Among Partisans",  
        x = "Party Affiliation", y = "Support for Trump") +  
  theme_bw() #black and white theme with grid lines
```



# Additional Resources

- ggplot2 cheat-sheet
- Hadley Wickham:  
<https://ggplot2.tidyverse.org>
- Healy, Kieran. (2018). Data Visualization: A Practical Introduction. Princeton University Press.  
Ch. 1: <http://assets.press.princeton.edu/chapters/s13826.pdf>
- Quick-R:  
<https://www.statmethods.net/index.html>

# Let's Practice!

