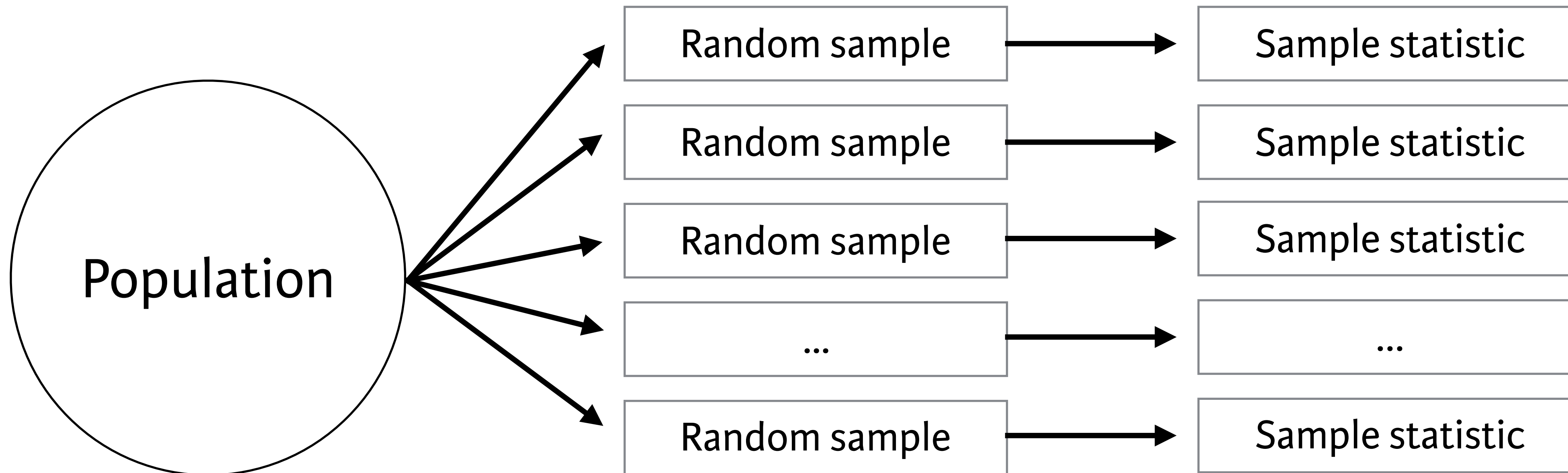# Precision of effect size estimates

Precision in modeling

- Confidence intervals are a representation of precision

- Take a computational perspective

  - Previous chapters: cross-validation

  - This chapter: bootstrapping

- Precision of effect sizes, prediction error, ...
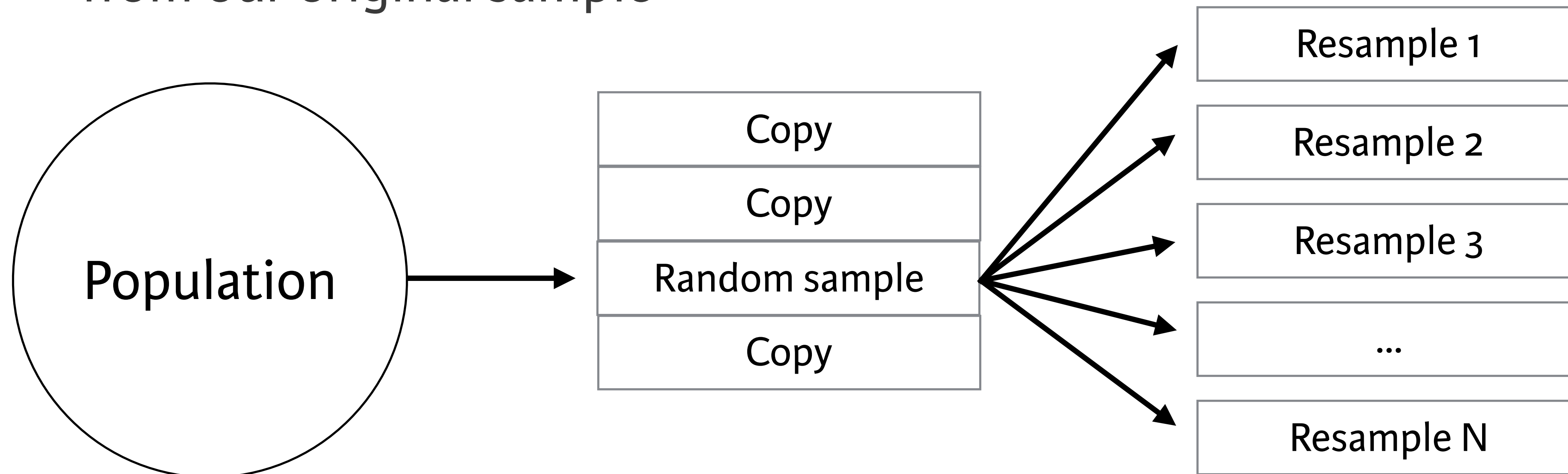
Framework for thinking about precision

Sampling distribution

Distribution of sample statistics
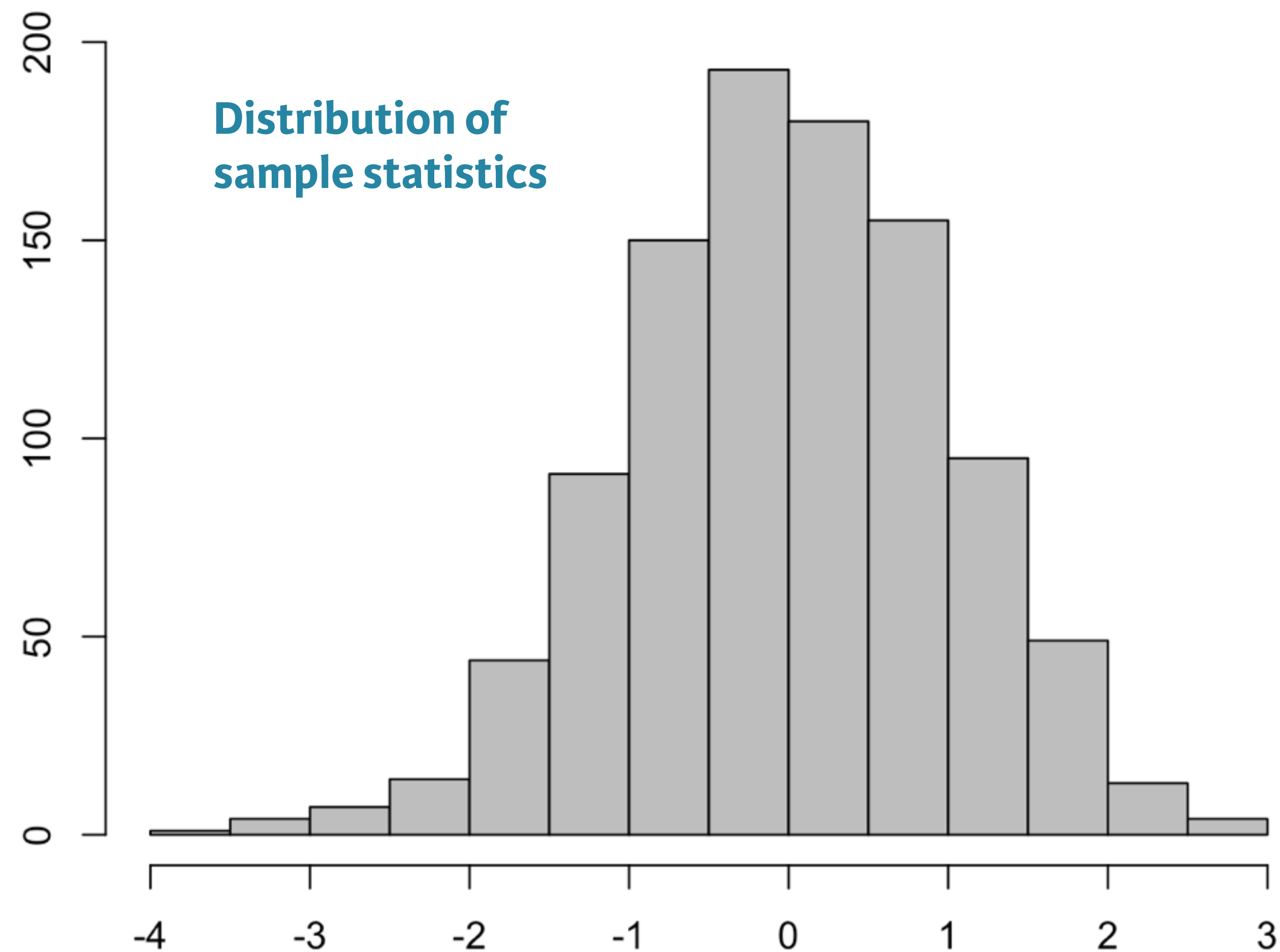
Bootstrapping

- Use a single random sample to simulate a situation where you have many random samples

- Use **resampling** to create "new" random samples from our original sample

Resampling in practice

```
> sample_of_statisticians
          name  yob
1        Efron 1938
2  Nightingale 1820
3          Box 1919
```

```
> library(mosaic)

# First resample
> sample(sample_of_statisticians, replace = TRUE)
      name  yob orig.id
1    Efron 1938       1
1.1  Efron 1938       1
3      Box 1919       3

# Second resample
> sample(sample_of_statisticians, replace = TRUE)
            name  yob orig.id
3            Box 1919       3
2     Nightingale 1820       2
3.1          Box 1919       3
```

Rinse and repeat

Resampling and effect size

```r
# Train a model on the data
> model <- lm(wage ~ age + exper + sex + sector, data = CPS85)
> effect_size(model, ~ age)
  slope age to:age exper sex sector
1  0.71  35     47    15   M   prof

# Train a model on resampled data
> model_resamp <- lm(wage ~ age + exper + sex + sector,
                     data = resample(CPS85))
> effect_size(model_resamp, ~ age)
  slope age to:age exper sex   sector
1  0.66  36     48    16   M clerical

# Each new resample is different from the previous ones
> model_resamp <- lm(wage ~ age + exper + sex + sector,
                     data = resample(CPS85))
> effect_size(model_resamp, ~ age)
  slope age to:age exper sex sector
1  0.87  34     46    15   M   prof
```
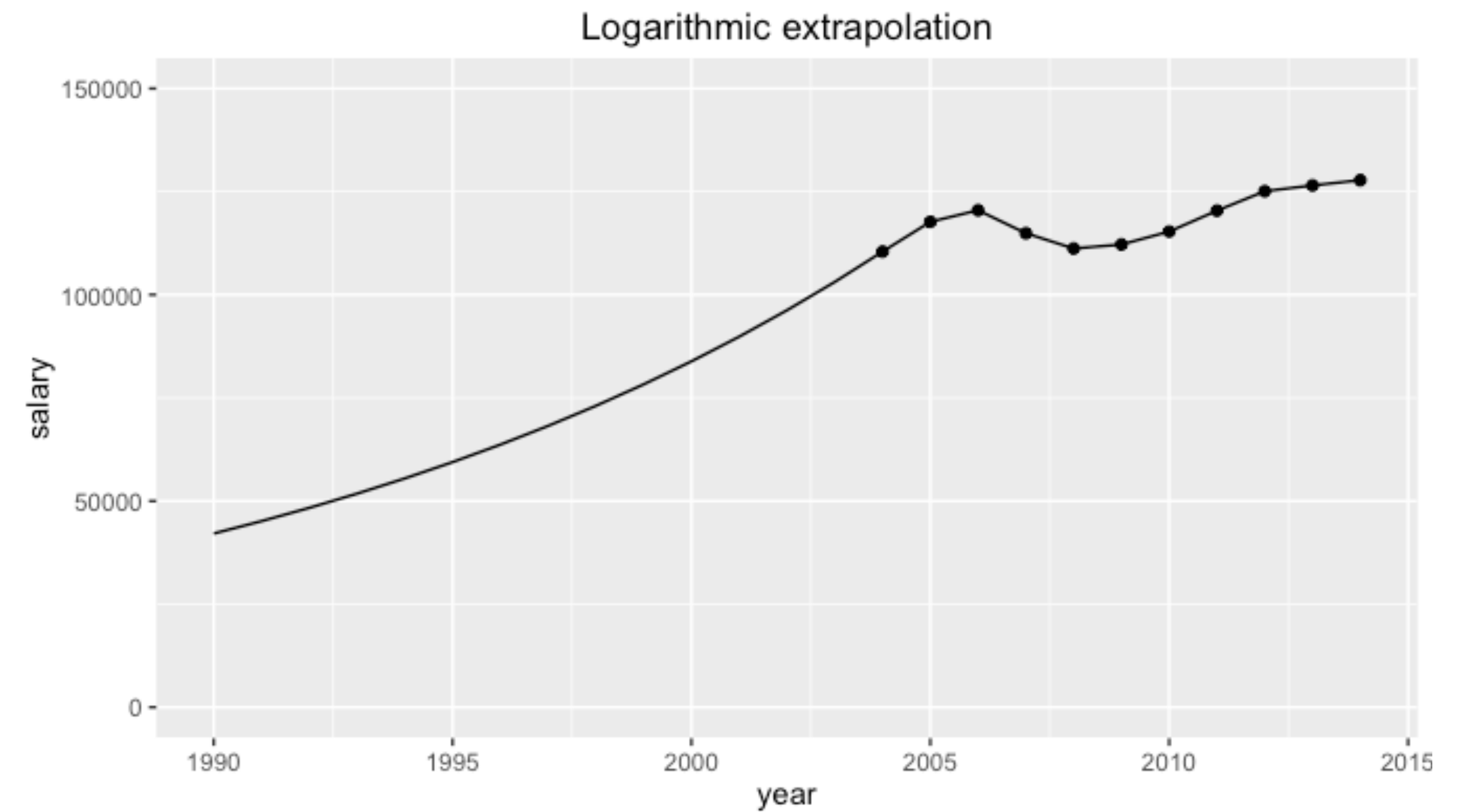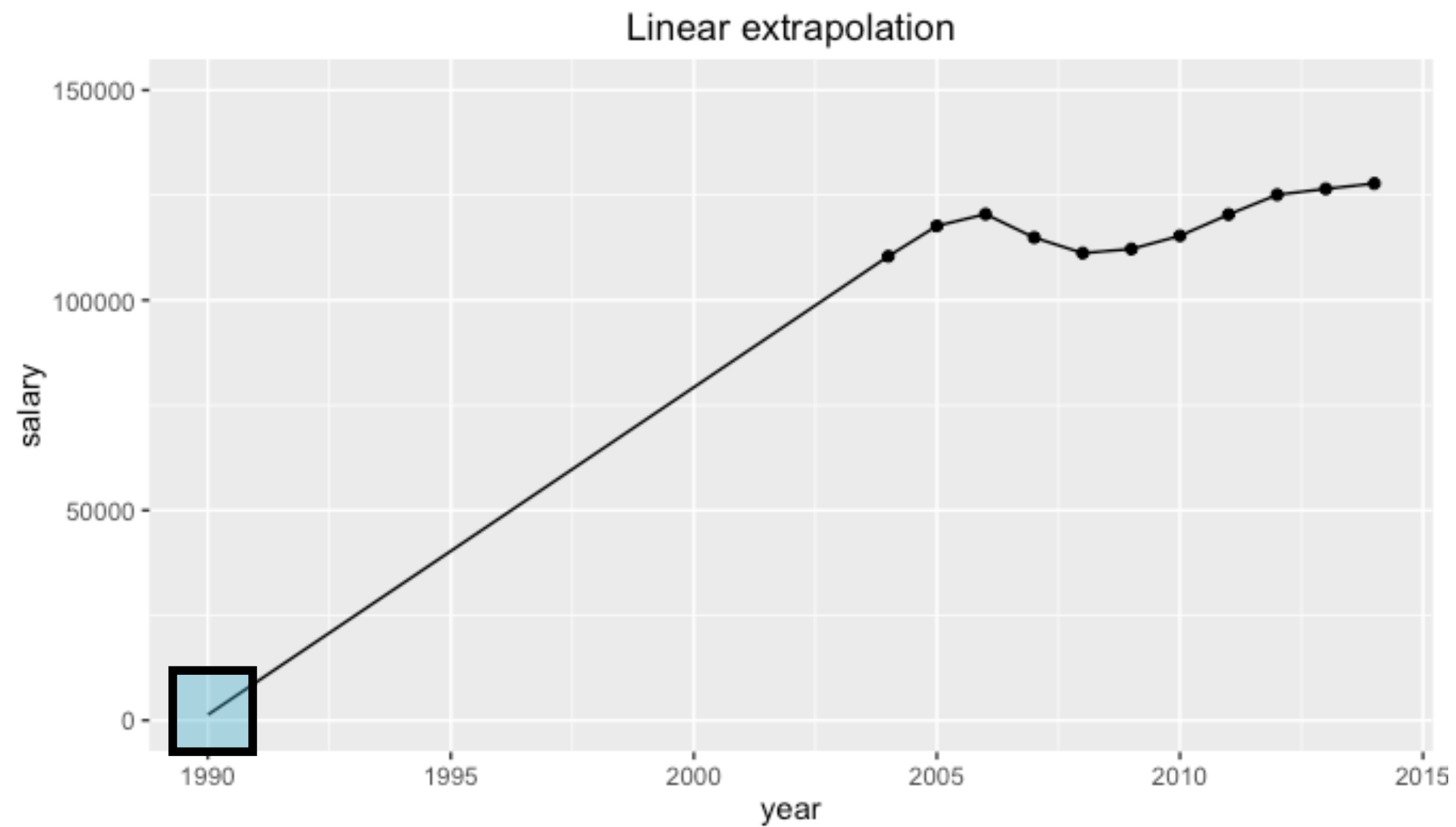
INTRODUCTION TO STATISTICAL MODELING

Let's practice!

INTRODUCTION TO STATISTICAL MODELING

Scales and transformations

Special situations

- Is the response 0 – 1?

- Is the response a count?

- Is the response cyclic (i.e. day of the week 1-7)?

- Is the response variable money, or another variable where change is proportional to current size?

Modeling salaries

Communicating about logarithms

```
# Effects of age on log wage
> CPS85$log_wage <- log(CPS85$wage)
> model <- lm(log_wage ~ age + exper + sex + sector,
              data = CPS85)


> effect_size(model, ~ age)
  slope age to:age exper sex sector
1 0.075  35    47    15   M   prof

# Convert to a proportion by exponentiation
> exp(0.075) - 1 # subtract 1 for interpretability
[1] 0.078   Wages go up 7.8 percent per year of age
```

The rank transform

The `rank()` function replaces each value with the position that value would occupy if the variable were sorted from least to most
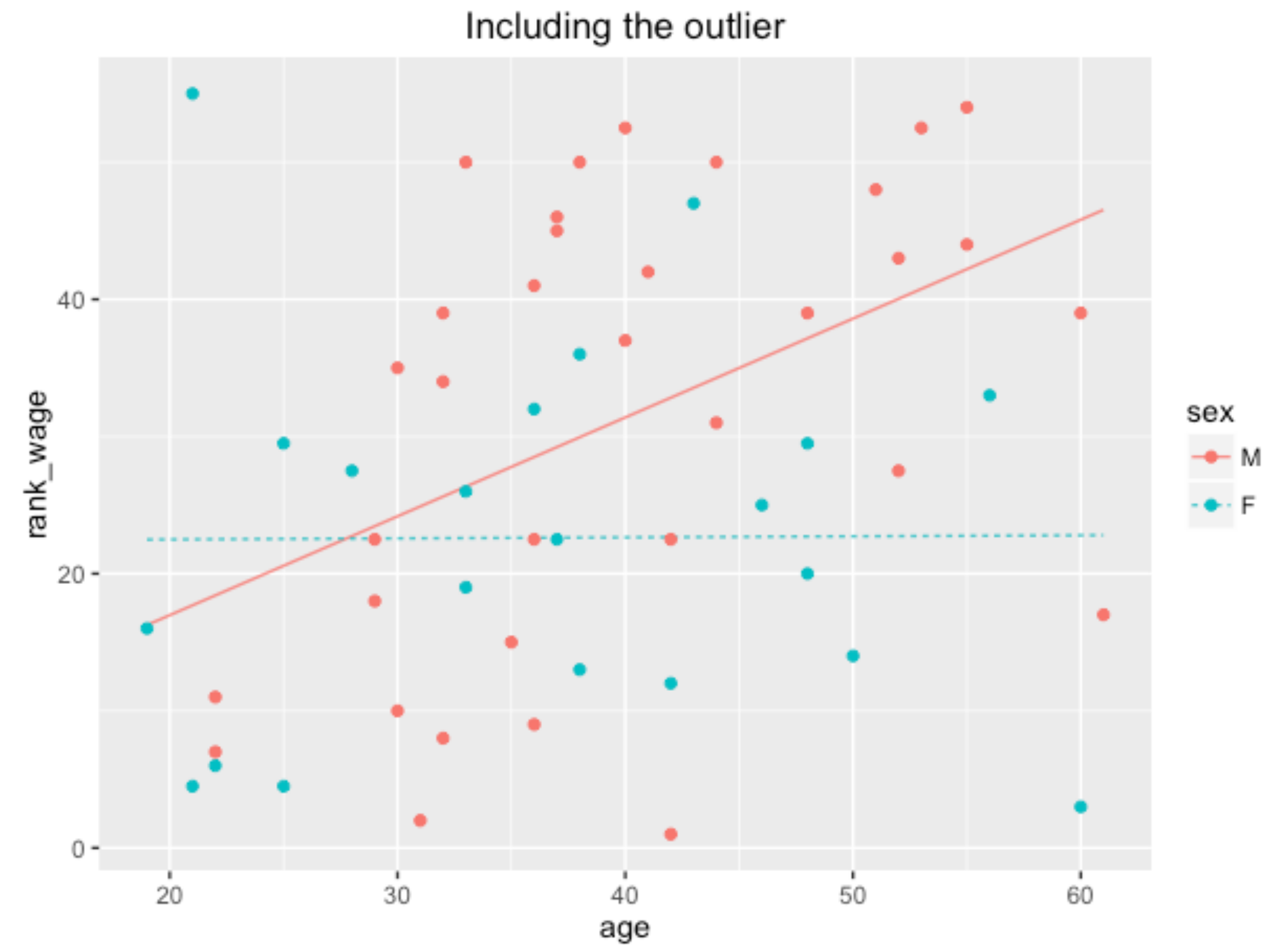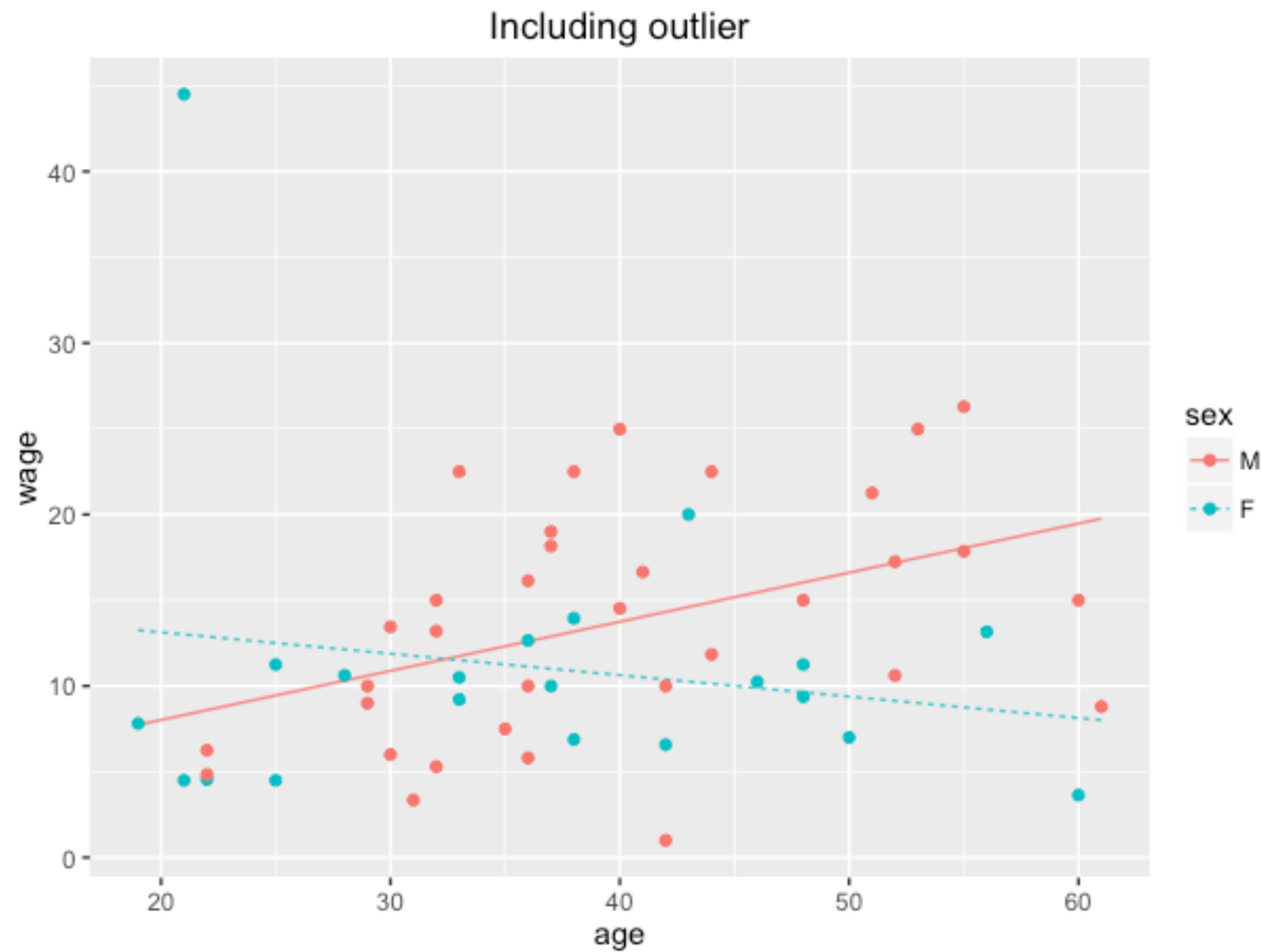
```
> example_vals <- c(2, 1000, 3, 8)
> example_vals
[1]    2 1000    3    8

> rank(example_vals)
[1] 1 4 2 3
```

Spotting an outlier

Applying rank transformation

INTRODUCTION TO STATISTICAL MODELING

Let's practice!