

Why Not Watch: Streaming A/B tests

Astrid de Geest

Effect on hours watched

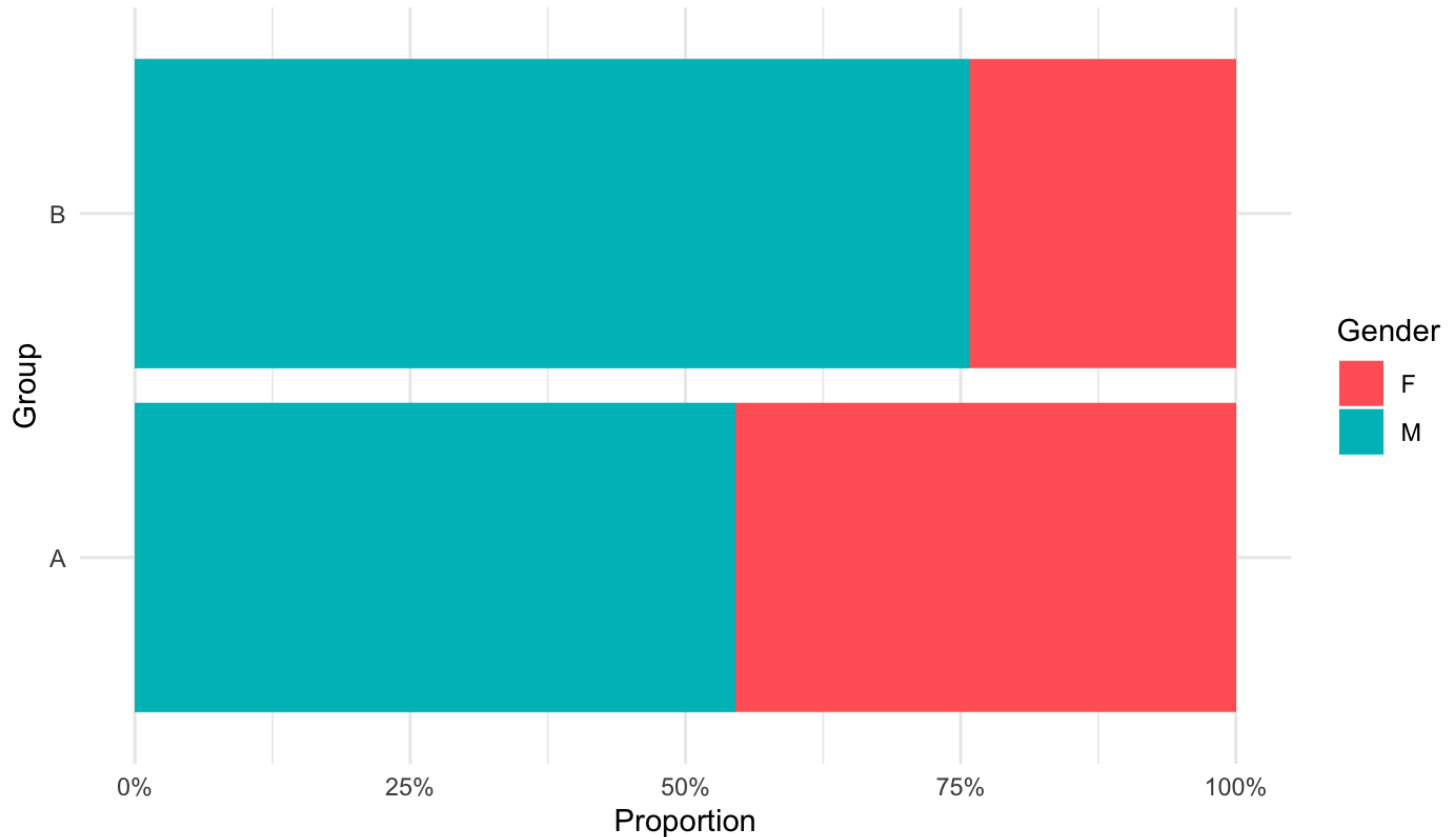
4:22

Group A

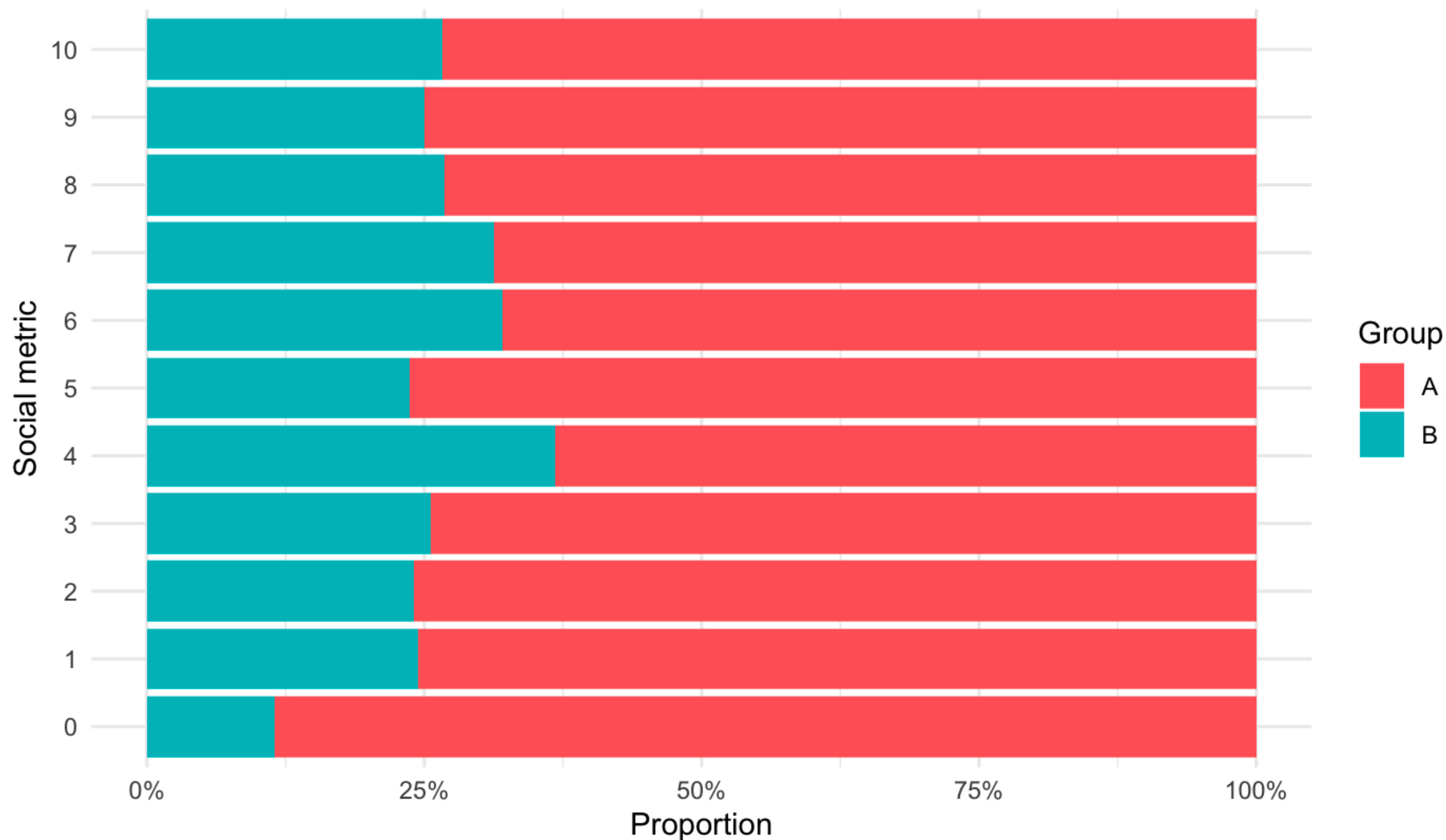
4:52

Group B

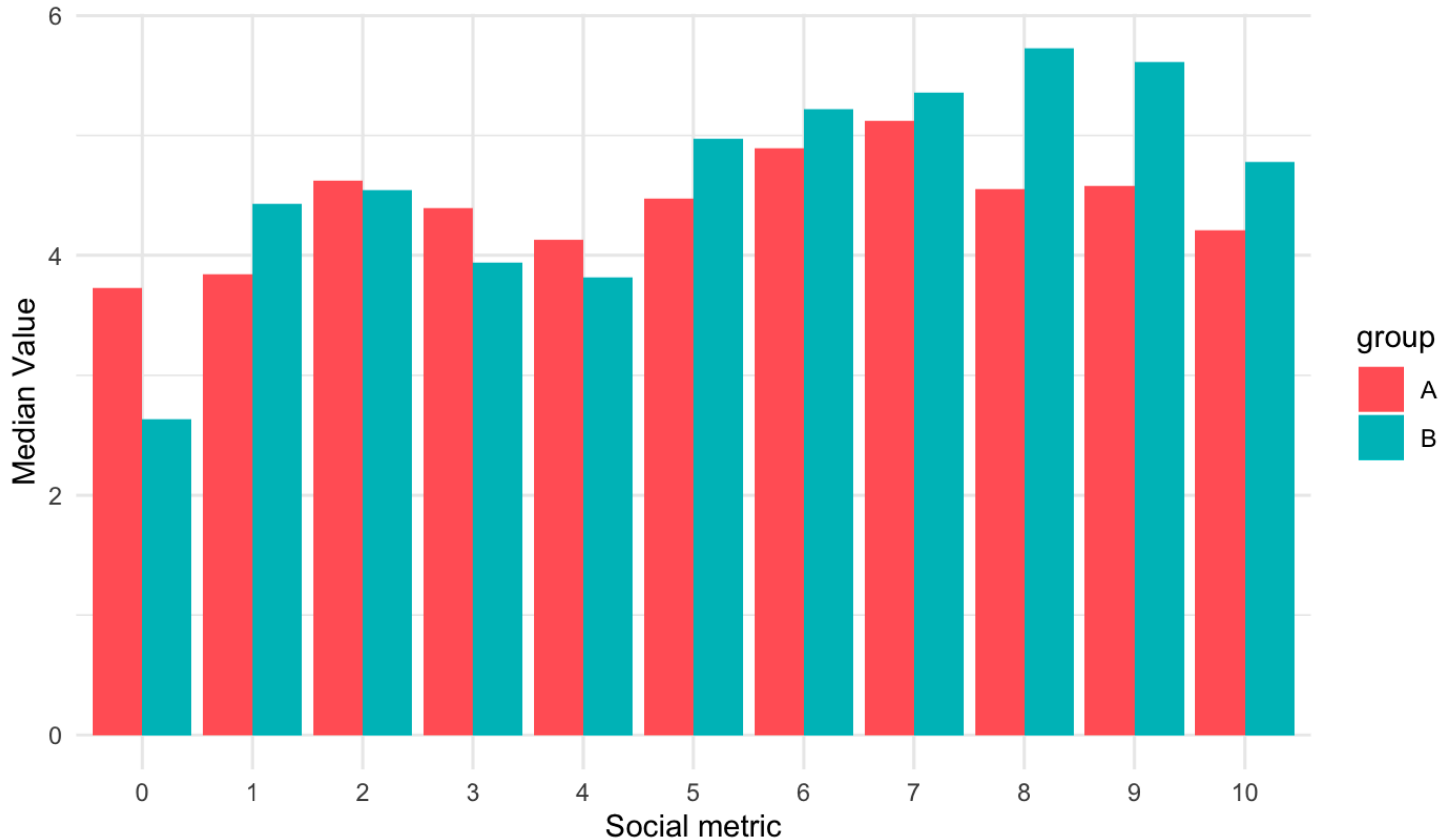
Gender Percentage of Groups A/B



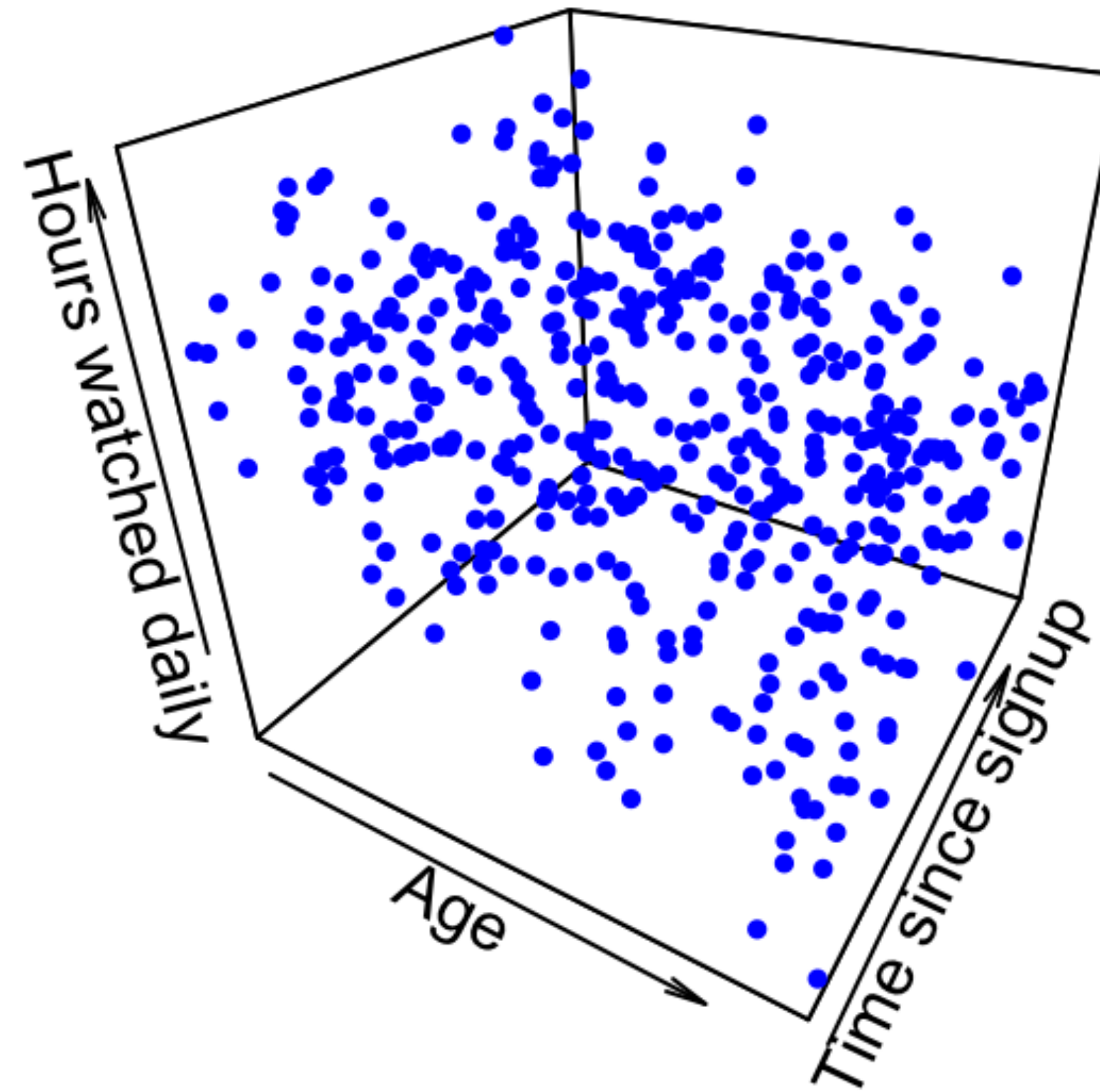
The proportion of each social metric in A/B testing



Median hours watched for A/B testing by social metric



Effect of age and time since last sign up on hours watched

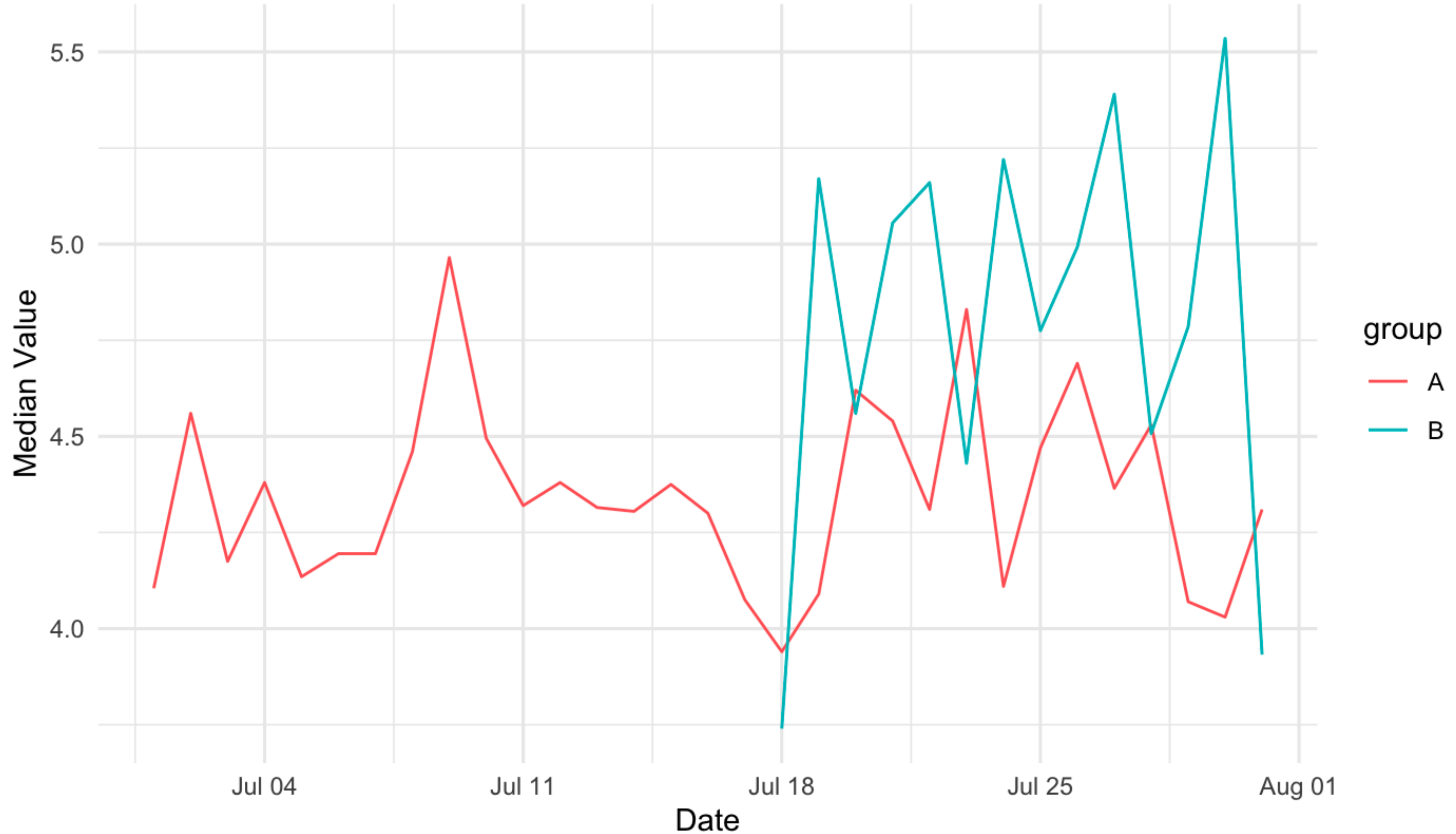


Age

Group A: 35 years

Group B: 39.5 years

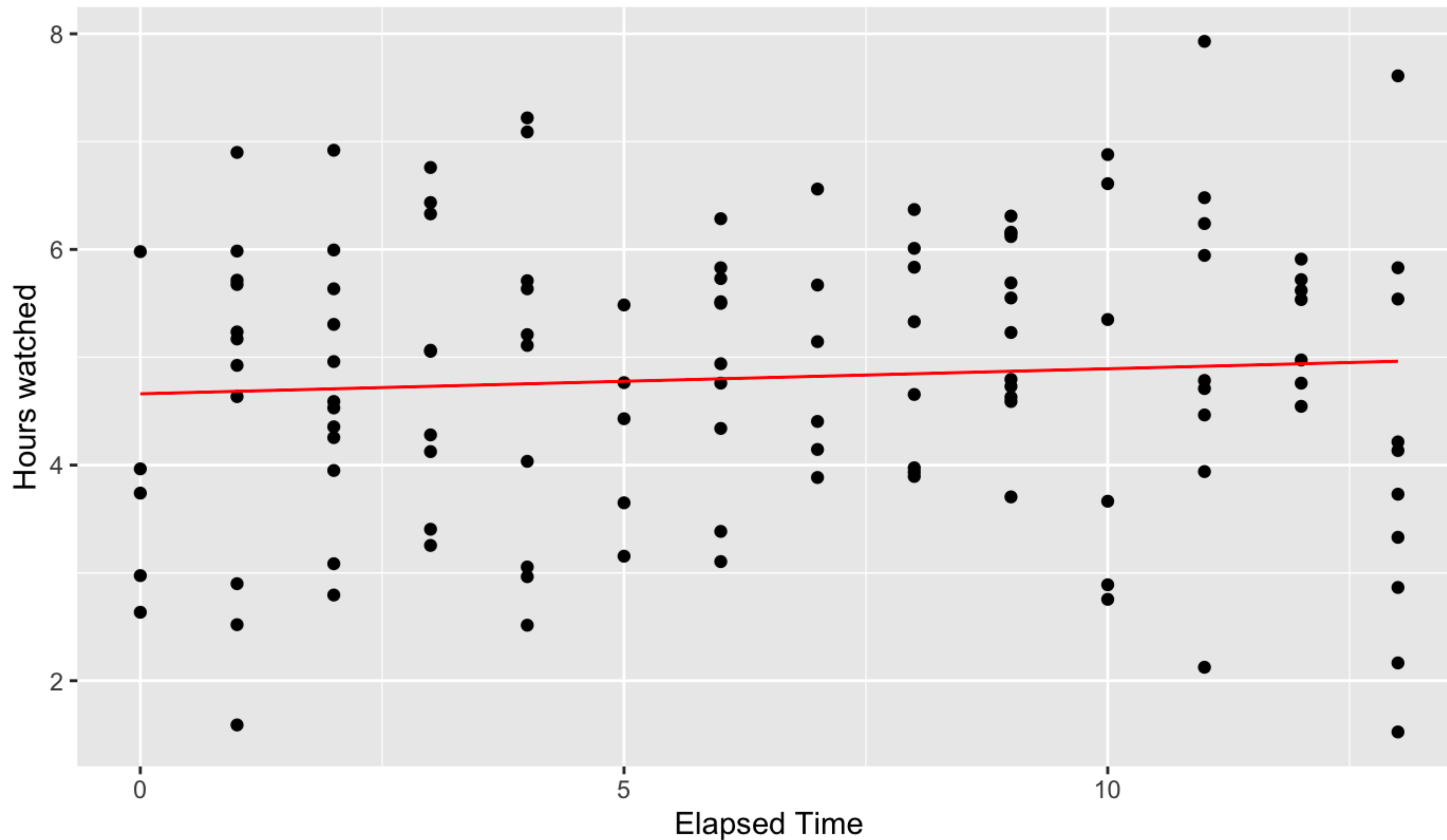
Average hours watched for A/B testing over dates



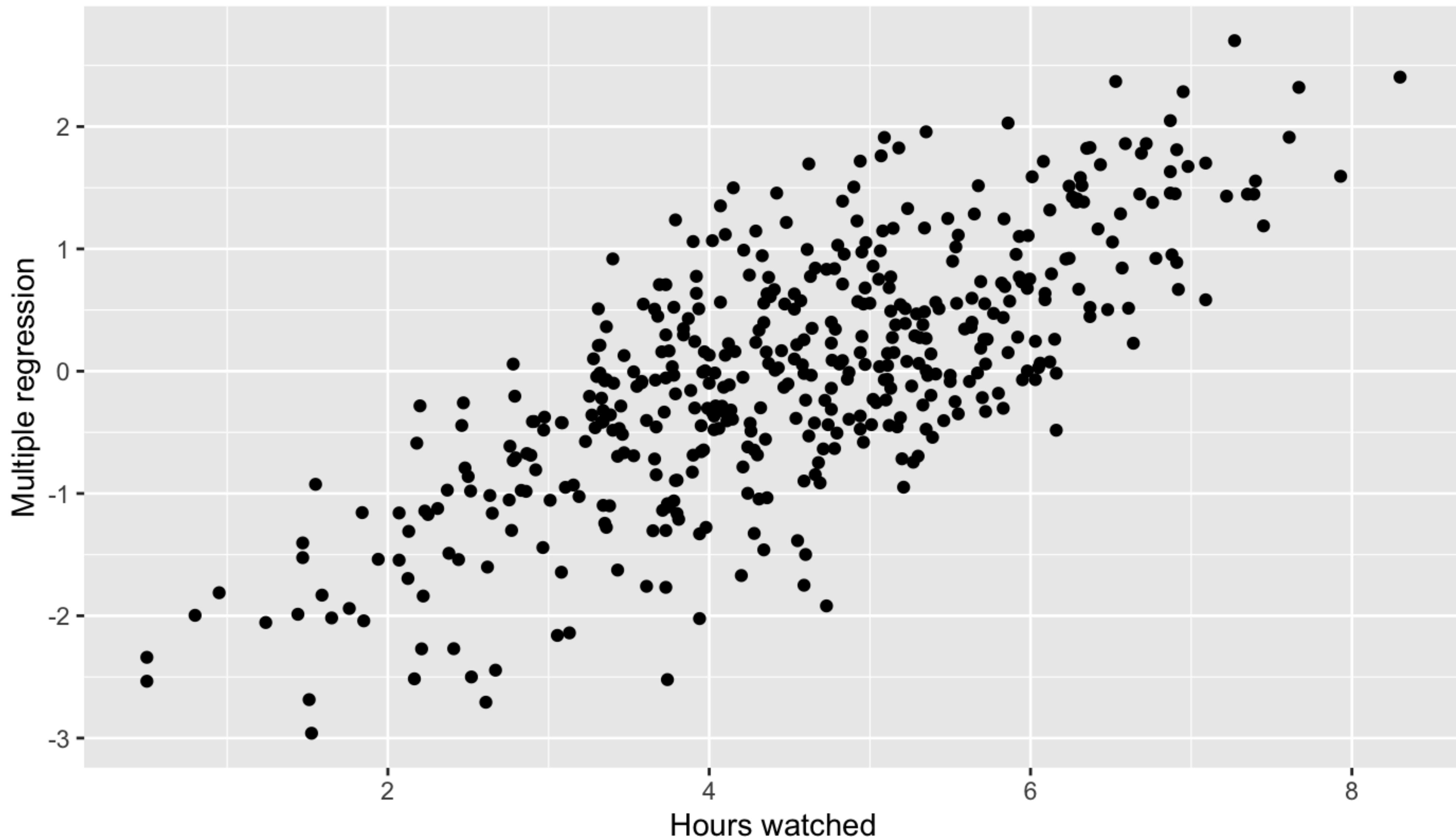
Bias in the data set

- ✦ Gender proportions
- ✦ Age difference with known impact on key variable
- ✦ Variation in social metric
- ✦ Variation in sampling for B's effect
- ✦ Samples from category A prior to start date of B

Time from recommendation engine and affect on hours watched



Multiple regression of hours watched with factors age, gender, A/B and social metric



A/B testing

- ✦ Calculate number required
- ✦ Random selection
- ✦ Check demographics
- ✦ Same timeframe for A/B group
- ✦ Consider average or multiple days of hours watched
- ✦ Account for novelty effect
- ✦ Intention with social metric/demographics