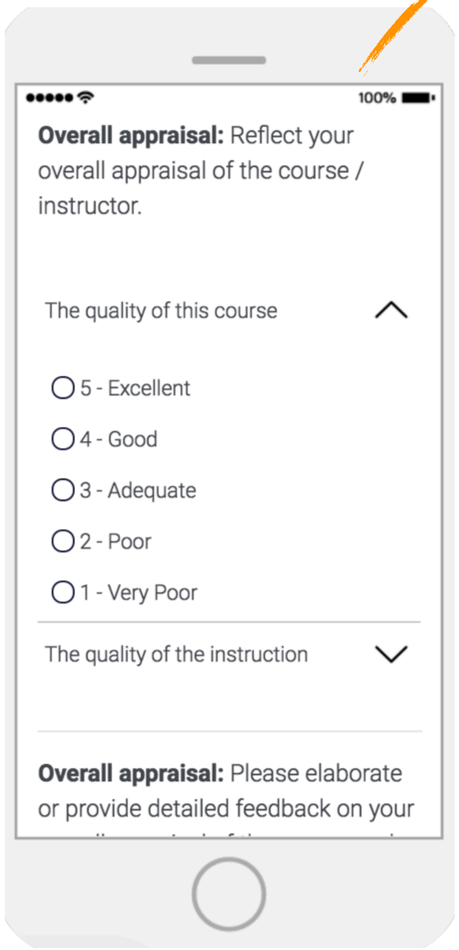
INTRODUCTION TO DATA

Beauty in the classroom

The data



score	rank	ethnicity	...	pic_color
4.7	tenure track	minority	...	color
4.1	tenure track	minority	...	color
3.9	tenure track	minority	...	color
...	...	...	...	...
4.1	tenure track	minority	...	color

Source: Hamermesh, Daniel S., and Amy Parker. "Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity." Economics of Education Review 24.4 (2005): 369-376.

INTRODUCTION TO DATA

Let’s practice!

INTRODUCTION TO DATA

Variables in the data

evals

```
> # Glimpse the data
> glimpse(evals)

Observations: 463
Variables: 21
$ score          <dbl> 4.7, 4.1, 3.9, 4.8, 4.6, 4.3...
$ rank           <fctr> tenure track, tenure track,...
$ ethnicity      <fctr> minority, minority, minorit...
$ gender         <fctr> female, female, female, fem...
$ language       <fctr> english, english, english, ...
$ age            <int> 36, 36, 36, 36, 59, 59, 59, ...
$ cls_perc_eval <dbl> 55.81, 68.80, 60.80, 62.60, ...
$ cls_did_eval   <int> 24, 86, 76, 77, 17, 35, 39, ...
$ cls_students   <int> 43, 125, 125, 123, 20, 40, 4...
$ cls_level      <fctr> upper, upper, upper, upper,...
$ cls_profs      <fctr> single, single, single, sin...
$ cls_credits    <fctr> multi credit, multi credit,...
```

evals (cont.)

```
> # Glimpse the data
> glimpse(evals)

...

$ bty_f1lower   <int> 5, 5, 5, 5, 4, 4, 4, 5, 5, 2...
$ bty_f1upper   <int> 7, 7, 7, 7, 4, 4, 4, 2, 2, 5...
$ bty_f2upper   <int> 6, 6, 6, 6, 2, 2, 2, 5, 5, 4...
$ bty_m1lower   <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 3...
$ bty_m1upper   <int> 4, 4, 4, 4, 3, 3, 3, 3, 3, 3...
$ bty_m2upper   <int> 6, 6, 6, 6, 3, 3, 3, 3, 3, 2...
$ bty_avg       <dbl> 5.000, 5.000, 5.000, 5.000, ...
$ pic_outfit    <fctr> not formal, not formal, not...
$ pic_color     <fctr> color, color, color, color,...
```

INTRODUCTION TO DATA

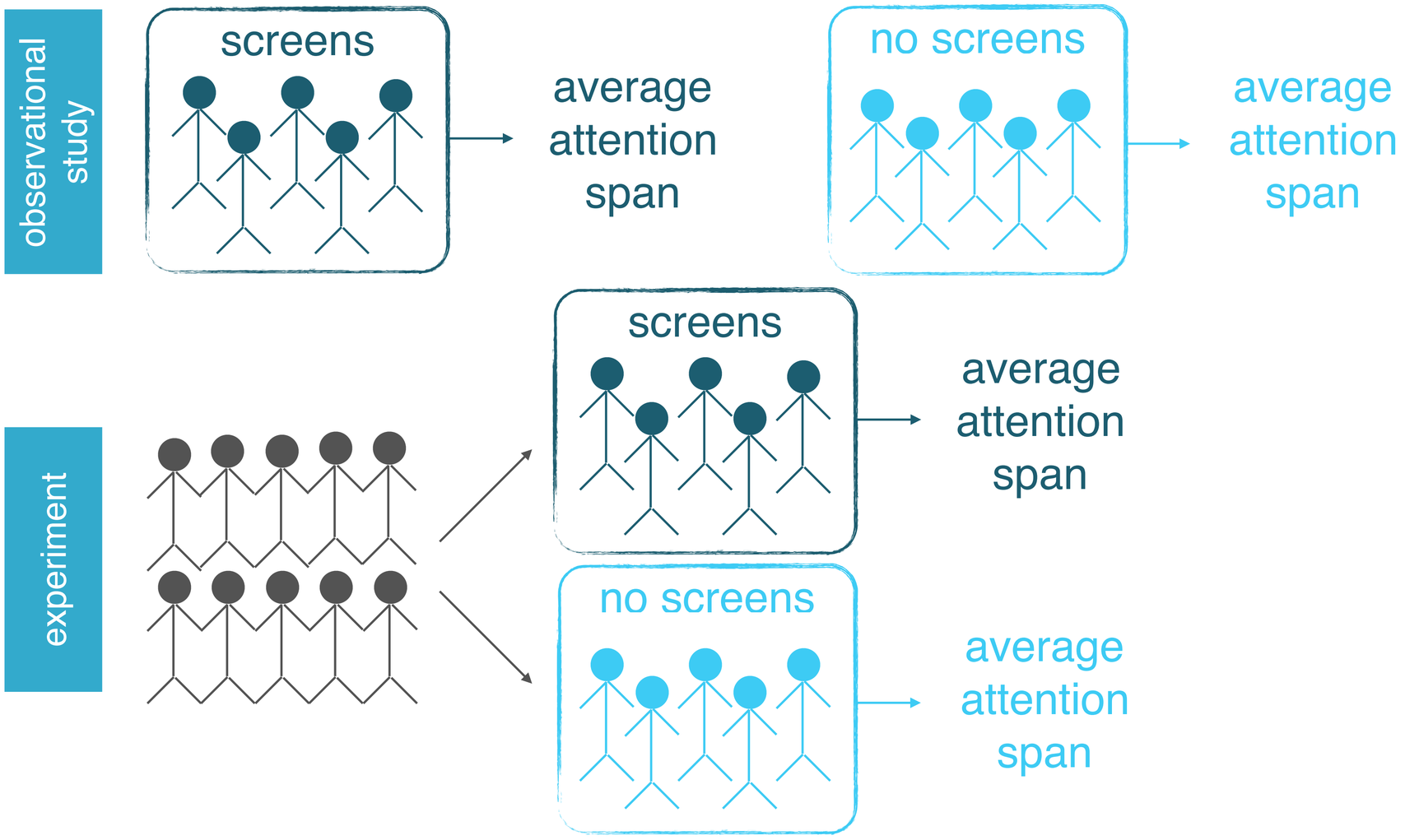
Let’s practice!

INTRODUCTION TO DATA

Congratulations!

Designing a study

Screens at bedtime and attention span



Viewing the structure of your data

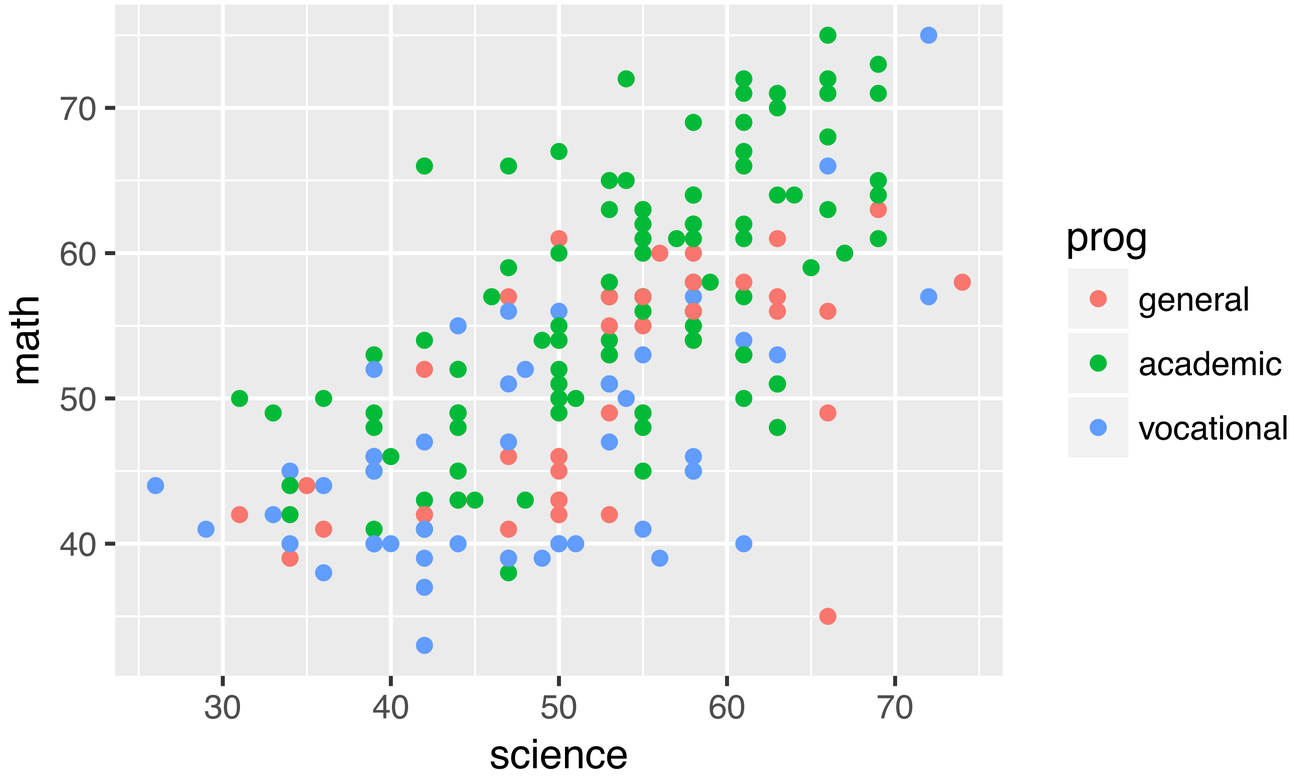
```
> # Load package
> library(dplyr)

> # View the structure of your data
> glimpse(hsb2)
Observations: 200
Variables: 11
$ id      <int> 70, 121, 86, 141, 172, 113, 50, 11, 84, 4...
$ gender  <chr> "male", "female", "male", "male", "male",...
$ race    <chr> "white", "white", "white", "white", "whit...
$ ses     <fctr> low, middle, high, high, middle, middle,...
$ schtyp  <fctr> public, public, public, public, public, ...
$ prog    <fctr> general, vocational, general, vocational...
$ read    <int> 57, 68, 44, 63, 47, 44, 50, 34, 63, 57, 6...
$ write   <int> 52, 59, 33, 44, 52, 52, 59, 46, 57, 55, 4...
$ math    <int> 41, 53, 54, 47, 57, 51, 42, 45, 54, 52, 5...
$ science <int> 47, 63, 58, 53, 53, 63, 53, 39, 58, 50, 5...
$ socst   <int> 57, 61, 31, 56, 61, 61, 61, 36, 51, 51, 6...
```

Data wrangling with dplyr

```
> # State distribution of SRS counties
> county_srs %>%
    group_by(state) %>%
    count()
# A tibble: 45 × 2
        state     n
       <fctr> <int>
1     Alabama     2
2      Alaska     1
3     Arizona     1
4    Arkansas     3
5  California     4
6    Colorado     2
7     Florida     3
8     Georgia     9
9       Idaho     2
10   Illinois     5
# ... with 35 more rows
```

Data visualization with ggplot2

```
> # Scatterplot of math vs. science scores, controlling for program
> ggplot(data = hsb2, aes(x = science, y = math, color = prog)) +
    geom_point()
```



INTRODUCTION TO DATA

Let’s practice!