Variance and Standard Deviation

Grinnell College

January 30, 2026

Review

Last time we ended with review of numerical summaries

- Measures of center
- Measures of dispersion

In particular, we considered two varieties: order and moment statistics

Variance

Today, we are going to take a closer look at variance:

- How is it defined
- Relationship between variance and standard deviation
- What is it used for?
  - Dispersion
  - Uncertainty
  - Prediction

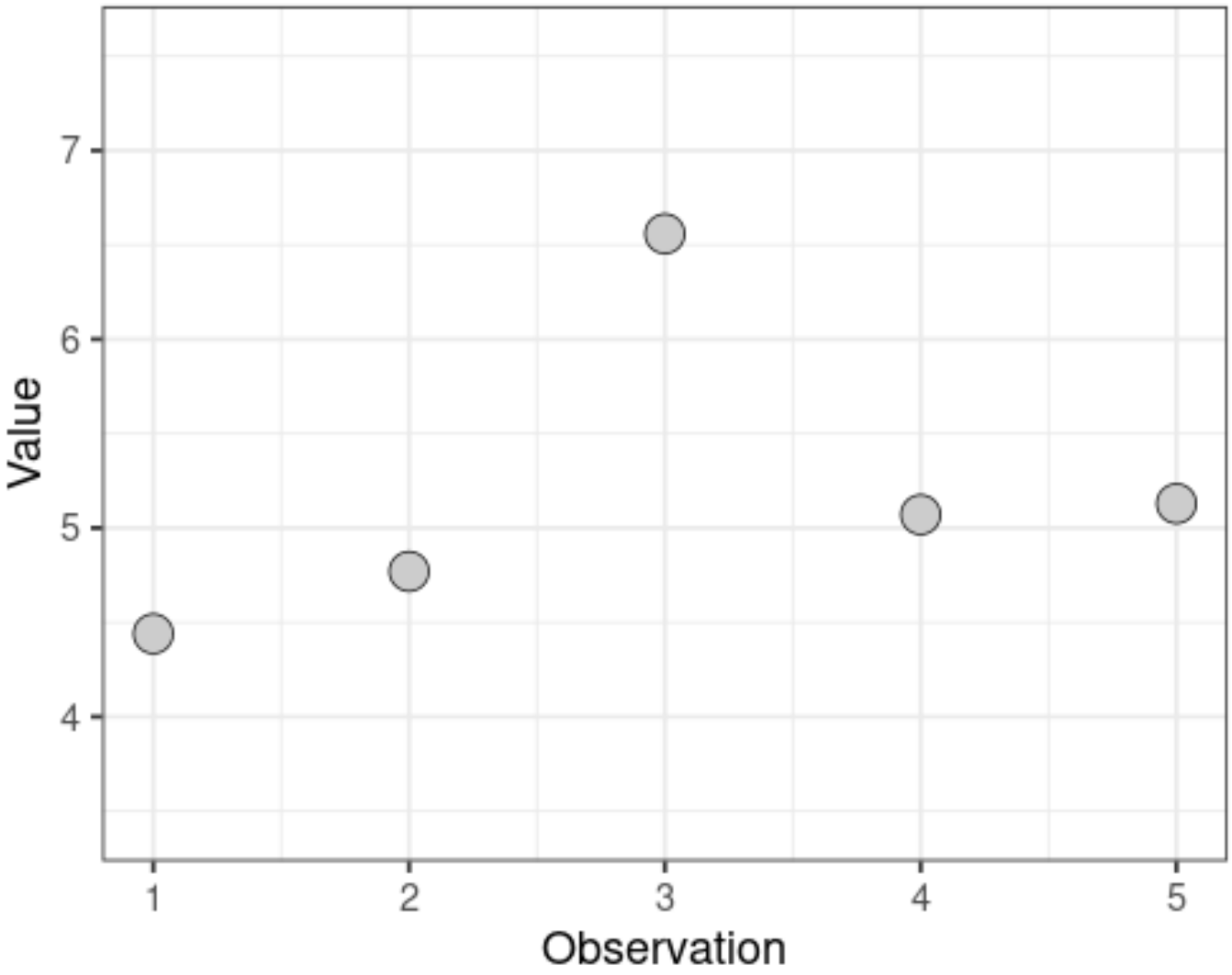
Most impactfully, the idea of variance is going to help us quantify statements such as: "this is the **best** guess we have"

Definitions

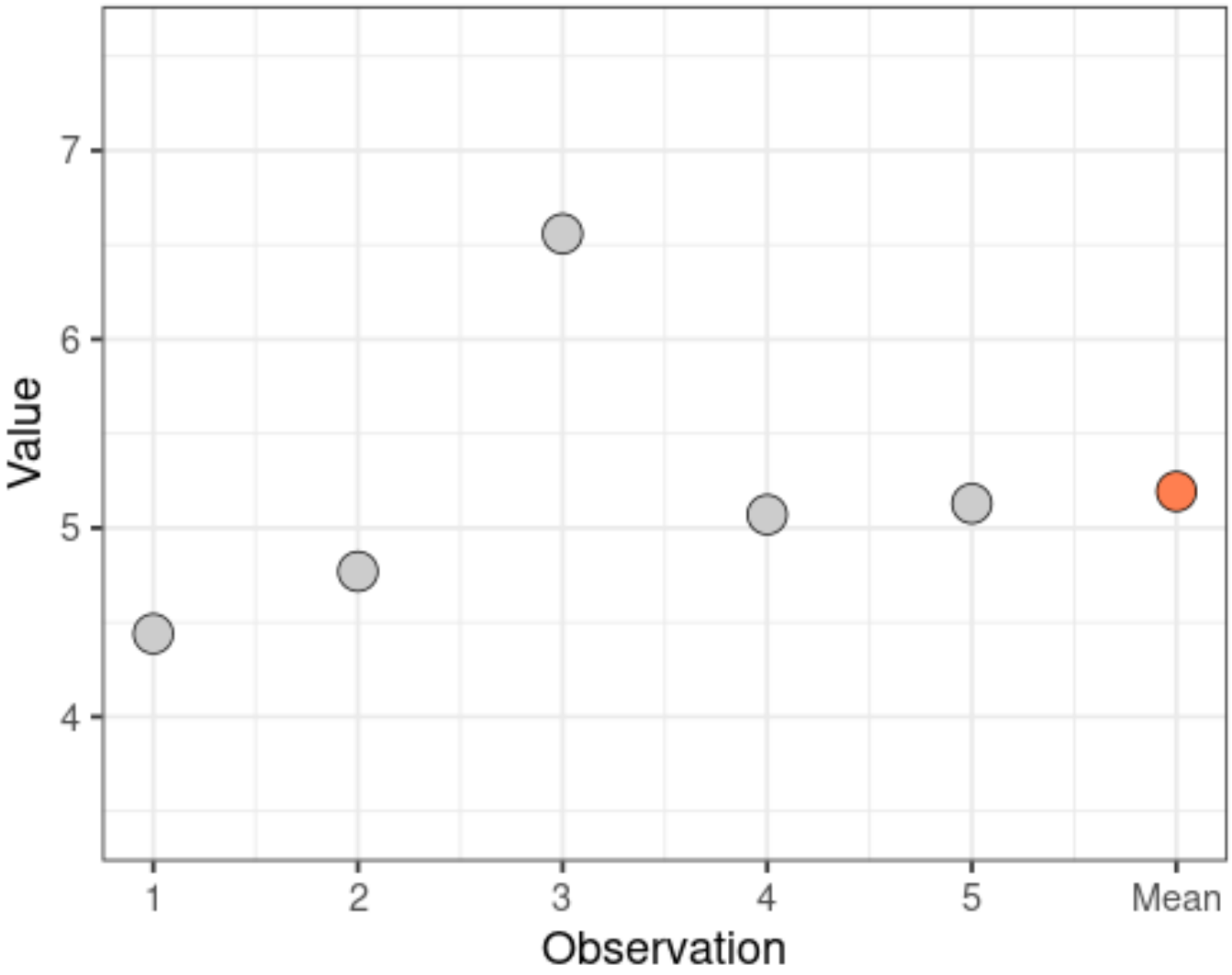
$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Just points

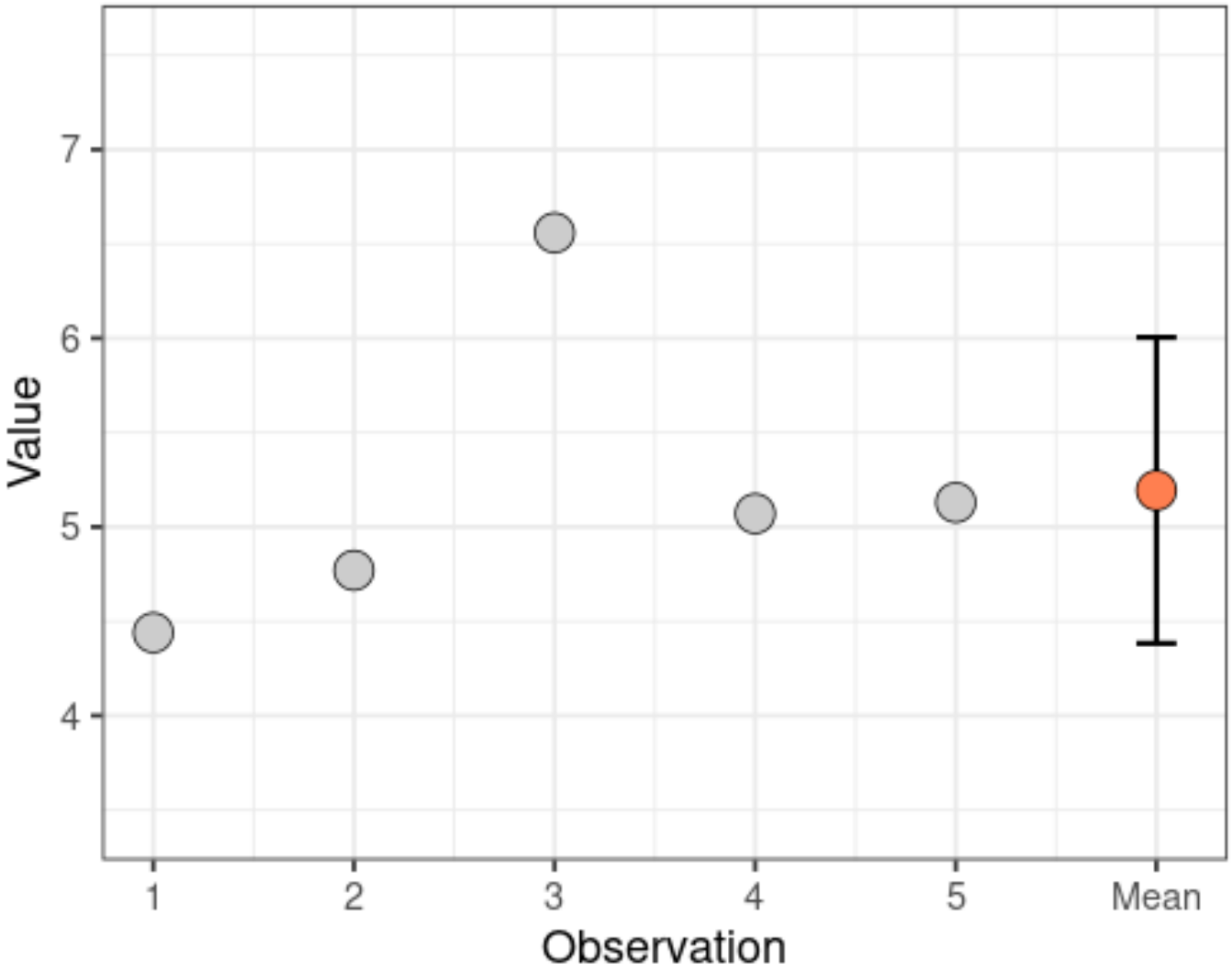


Just points



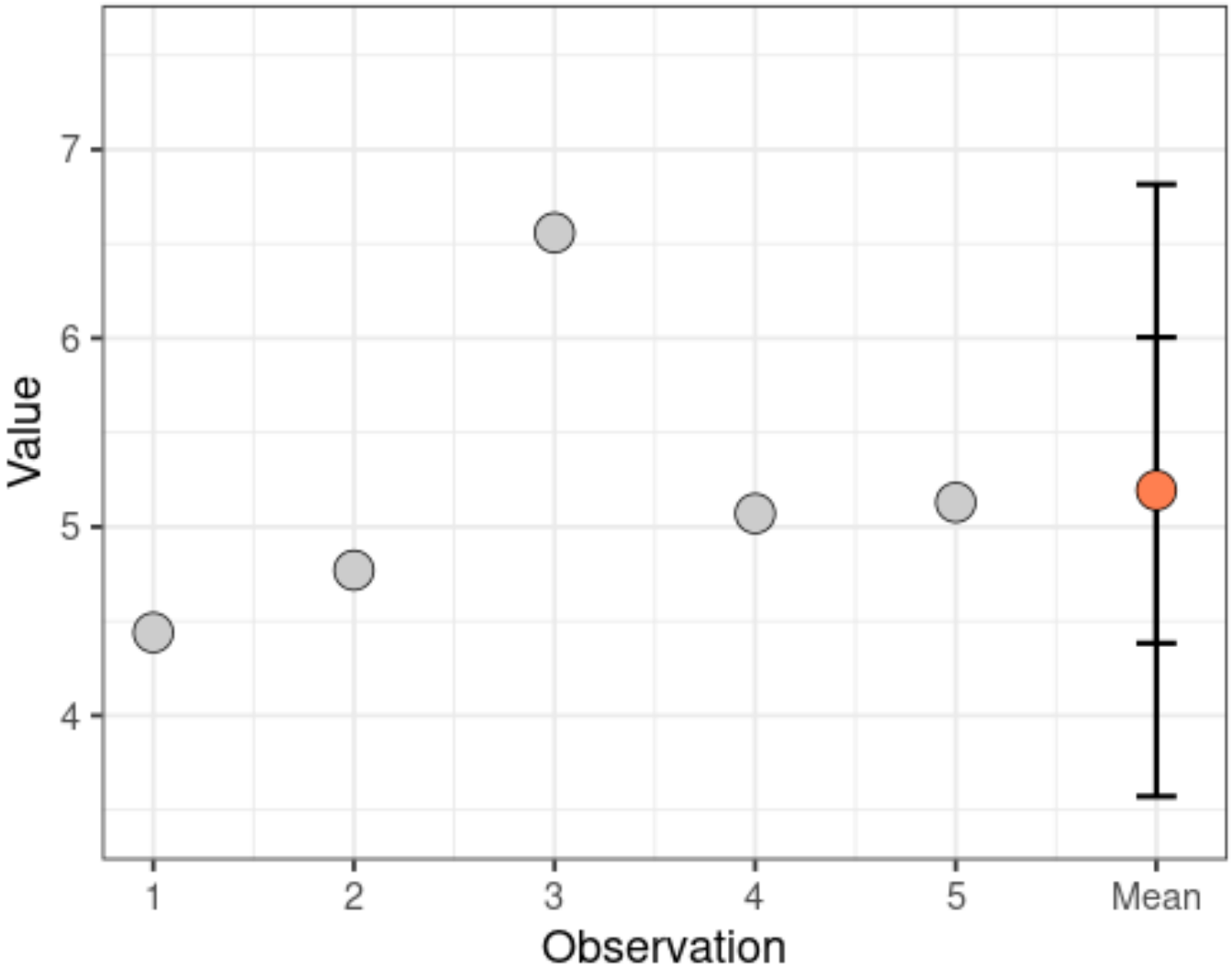
Just points

Here $n = 5$, $\overline{x} = 5.19$ and $\hat{\sigma} = 0.81$



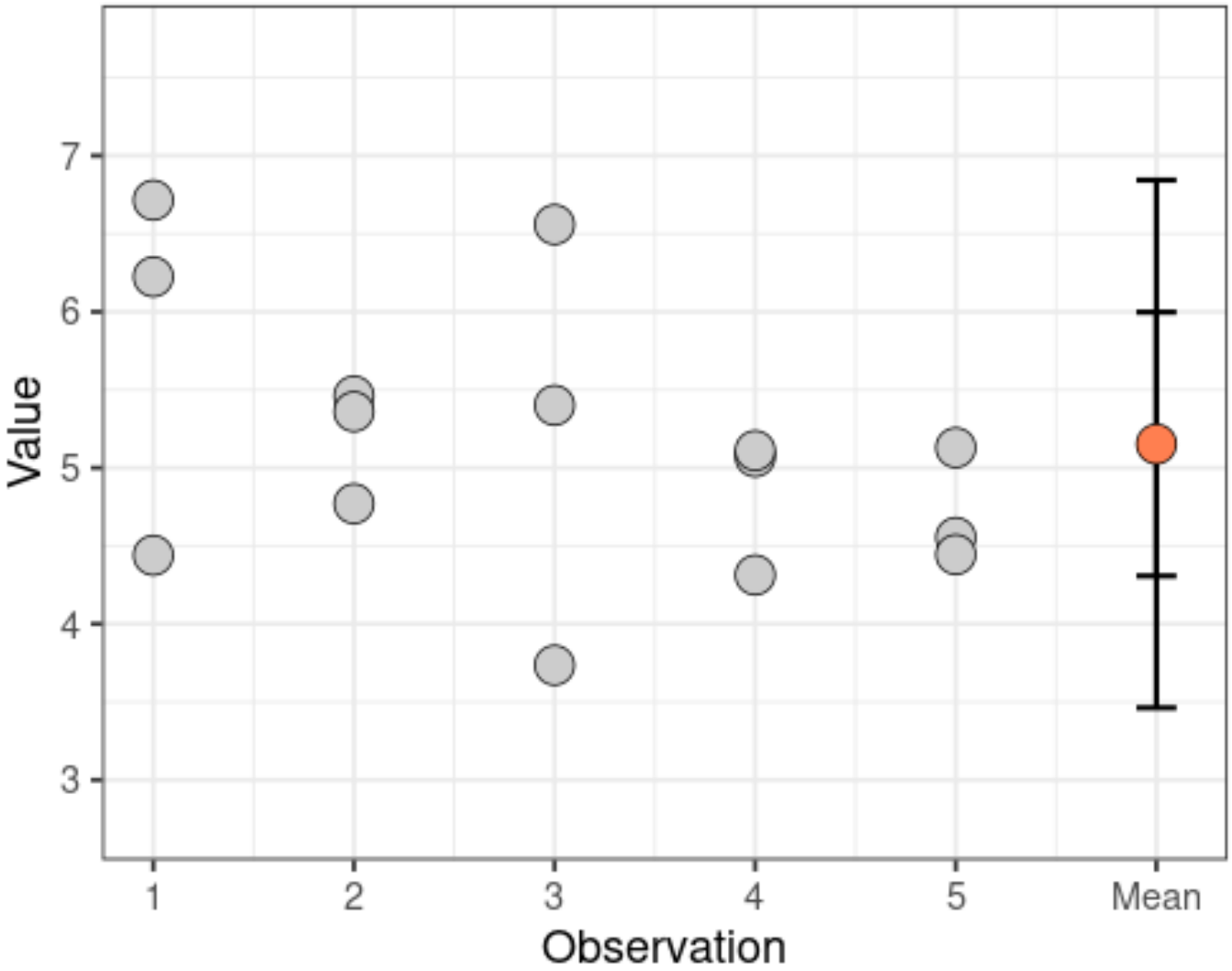
Just points

Here $n = 5$, $\overline{x} = 5.19$ and $\hat{\sigma} = 0.81$



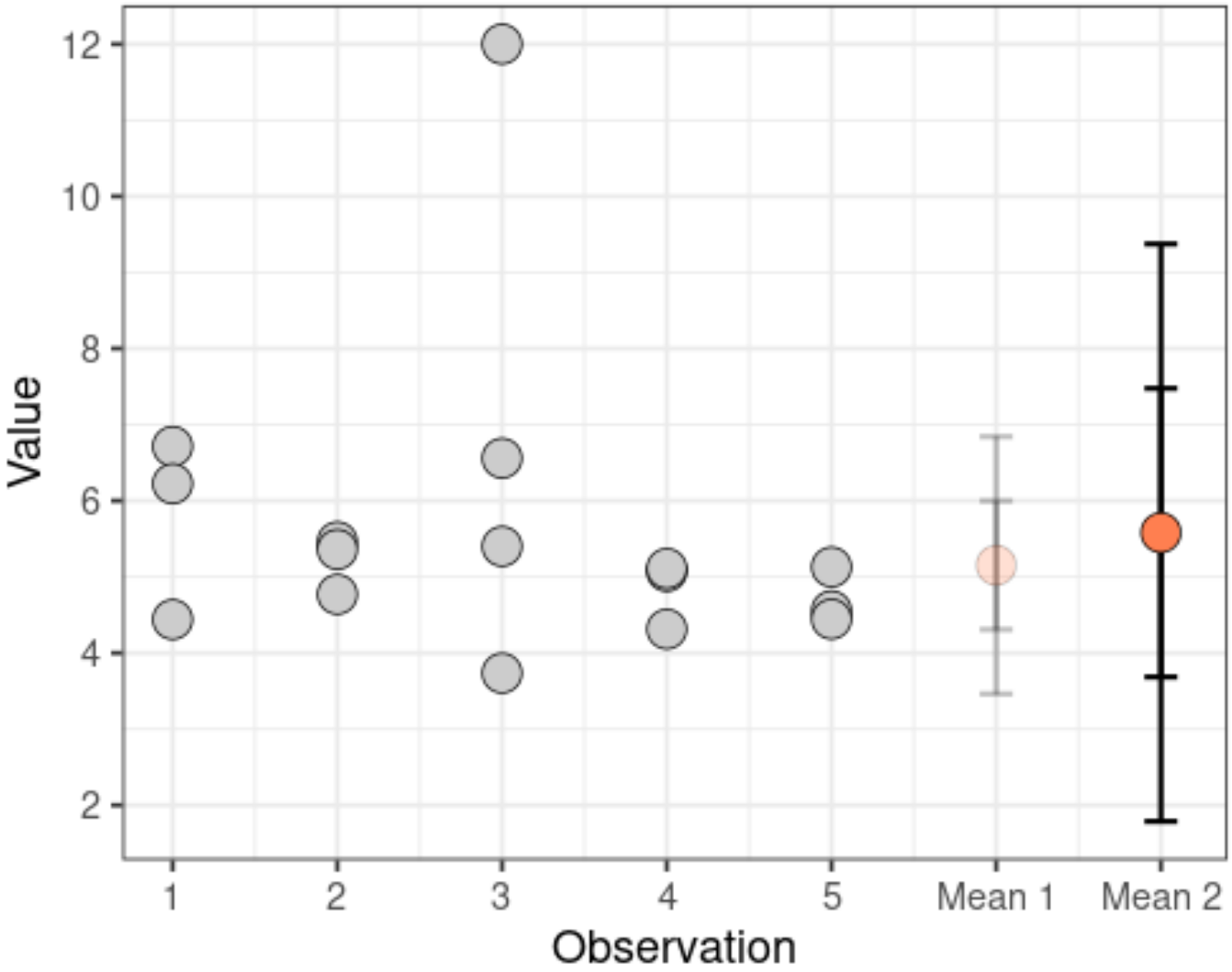
Just points

Note that it is not impacted by the number of observations. Here $n = 10$, $\overline{x} = 5.15$ and $\hat{\sigma} = 0.83$



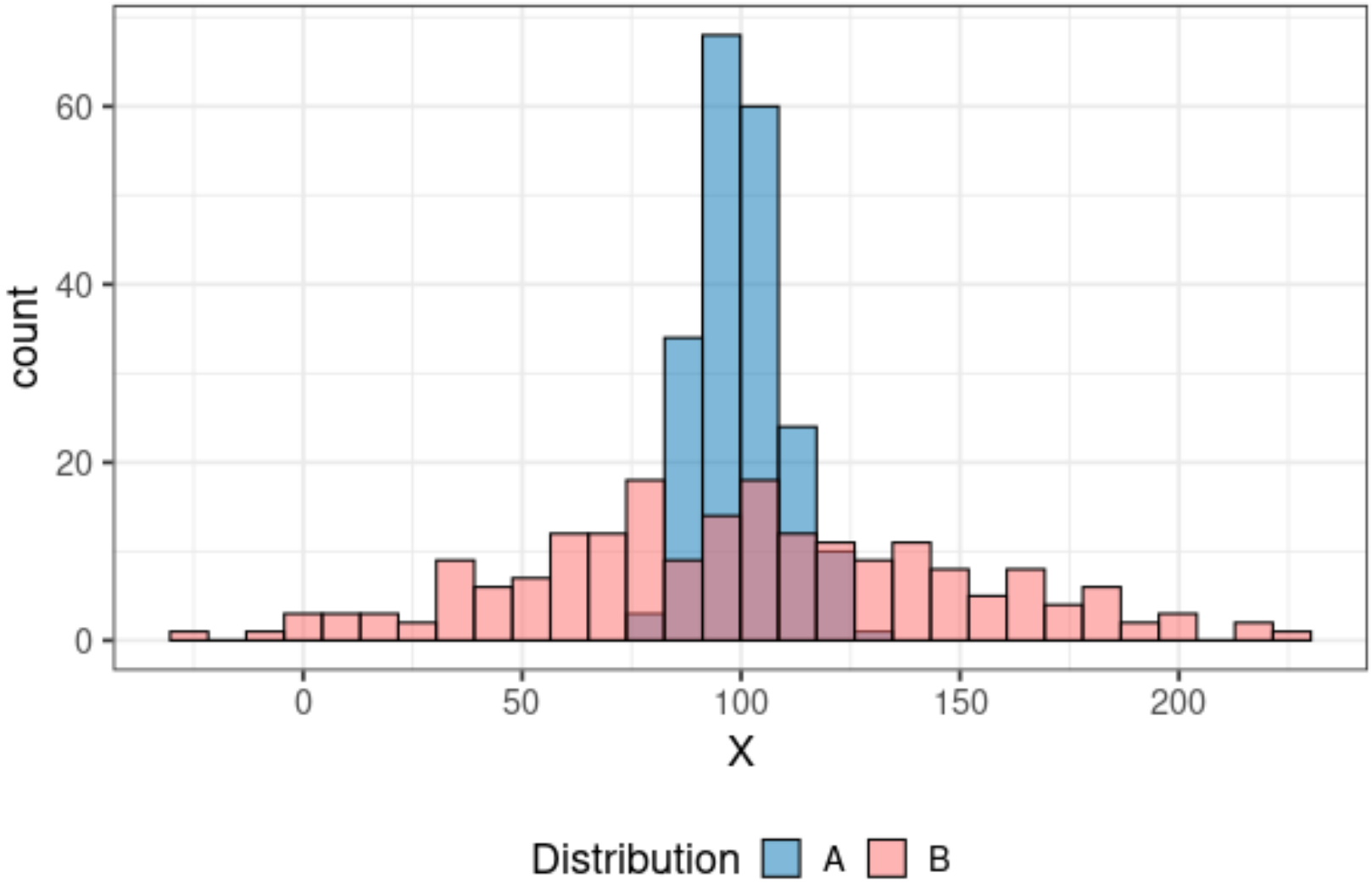
Outlier

Now $n = 11$, $\overline{x} = 5.6$ and $\hat{\sigma} = 1.9$

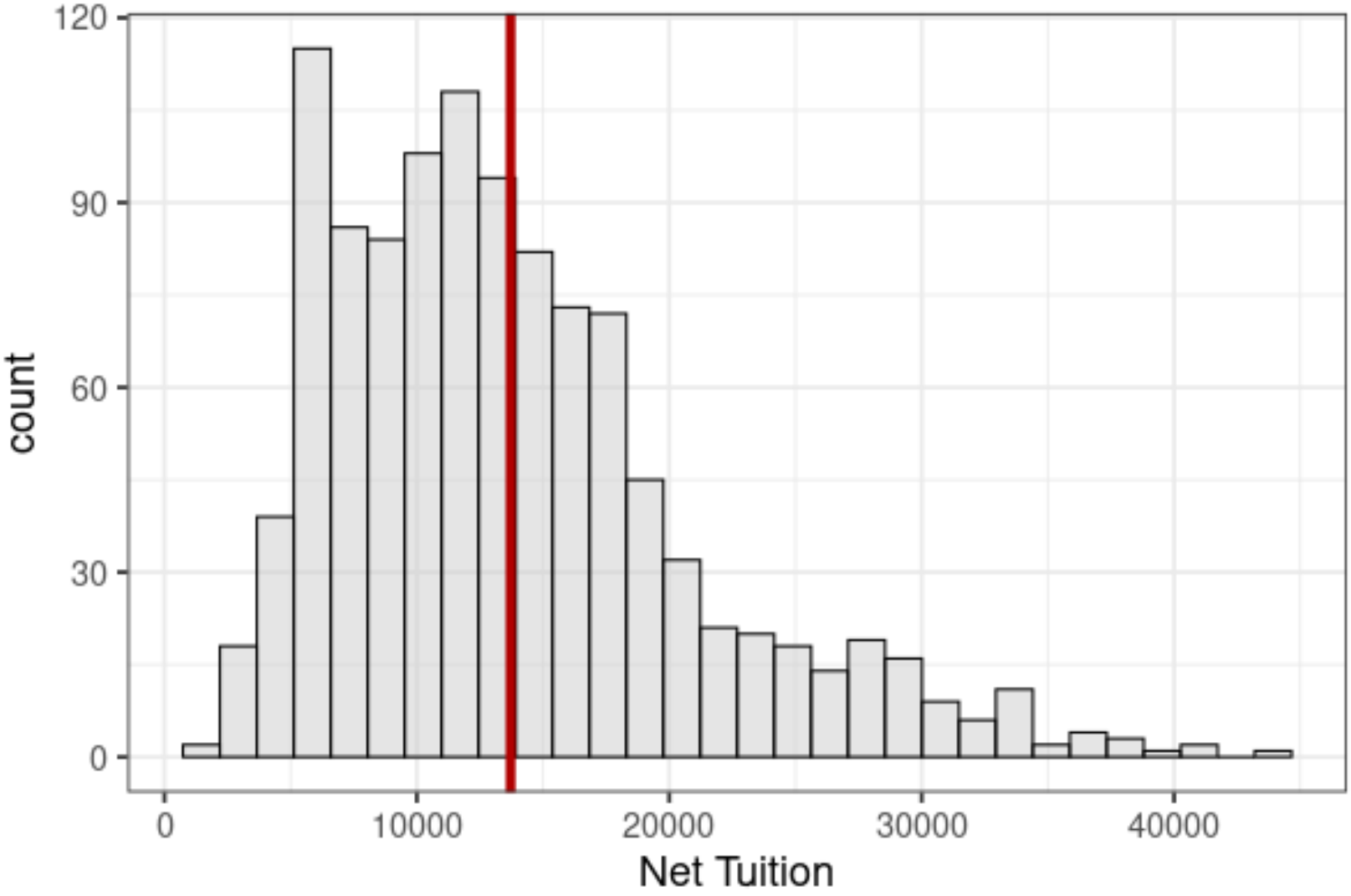


Dispersion

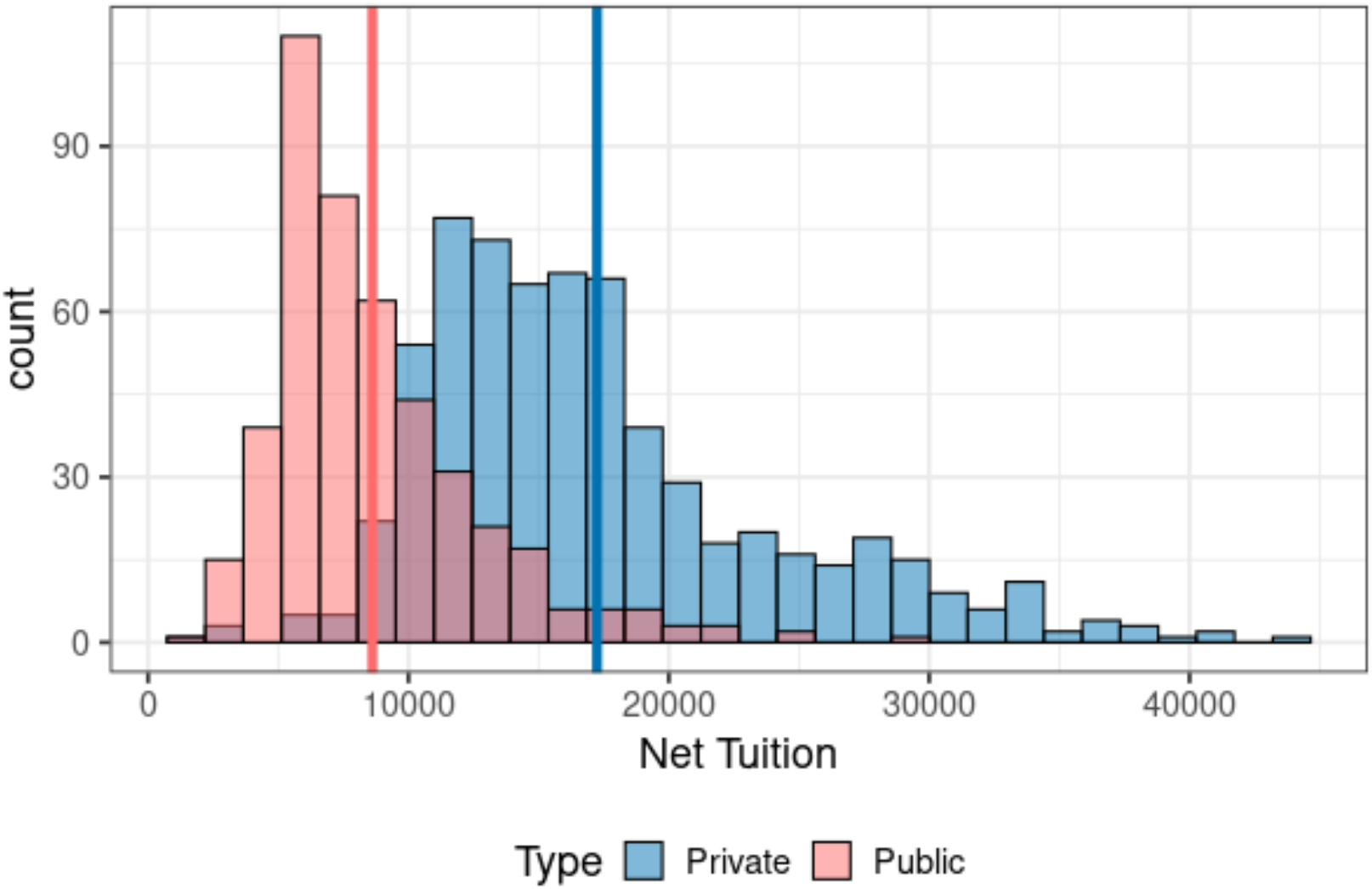
Both of these have $\mu = 100$



Better Centers?



Better Centers?



Main Takeaways

1. Variance and standard deviation are measures of dispersion
2. Standard deviation represents "average distance" (kinda)
3. Can be used to quantify outliers
4. Used to represent uncertainty or error in prediction
5. Ain't gonna do it by hand