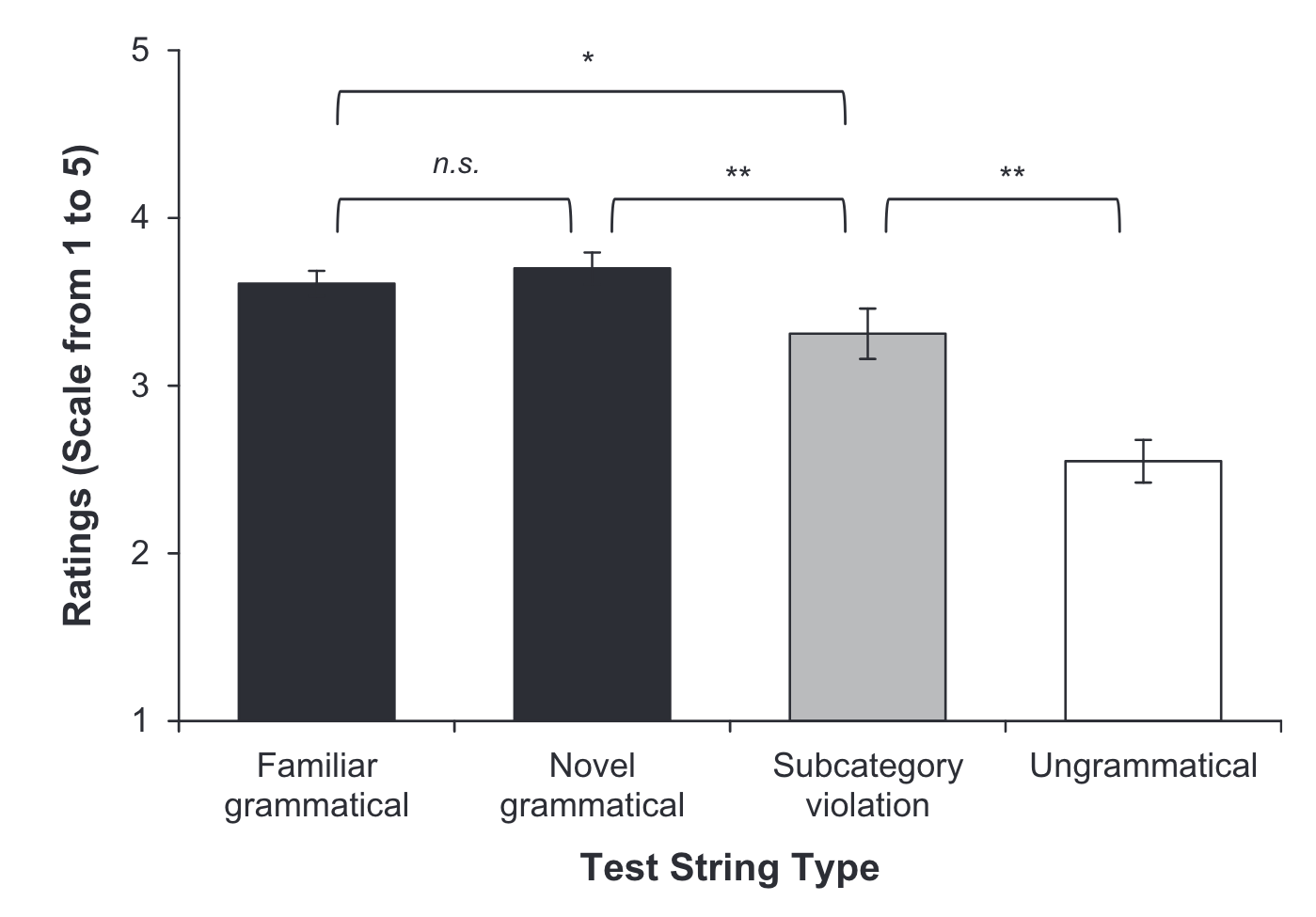
Three common mistakes in statistics   
and how to avoid them

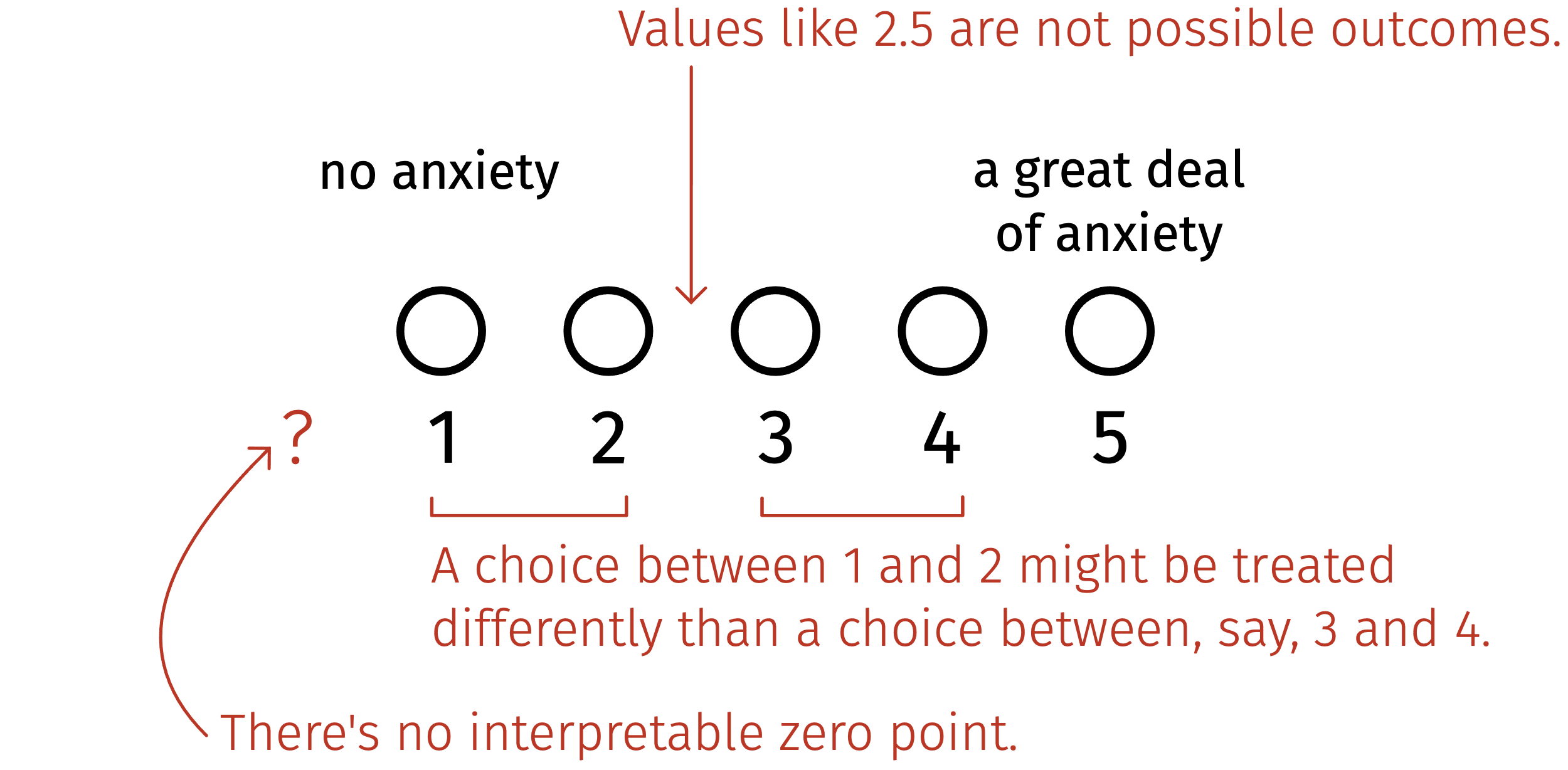
Elizabeth Pankratz, 26 March 2025

# Something you won’t be able to unsee

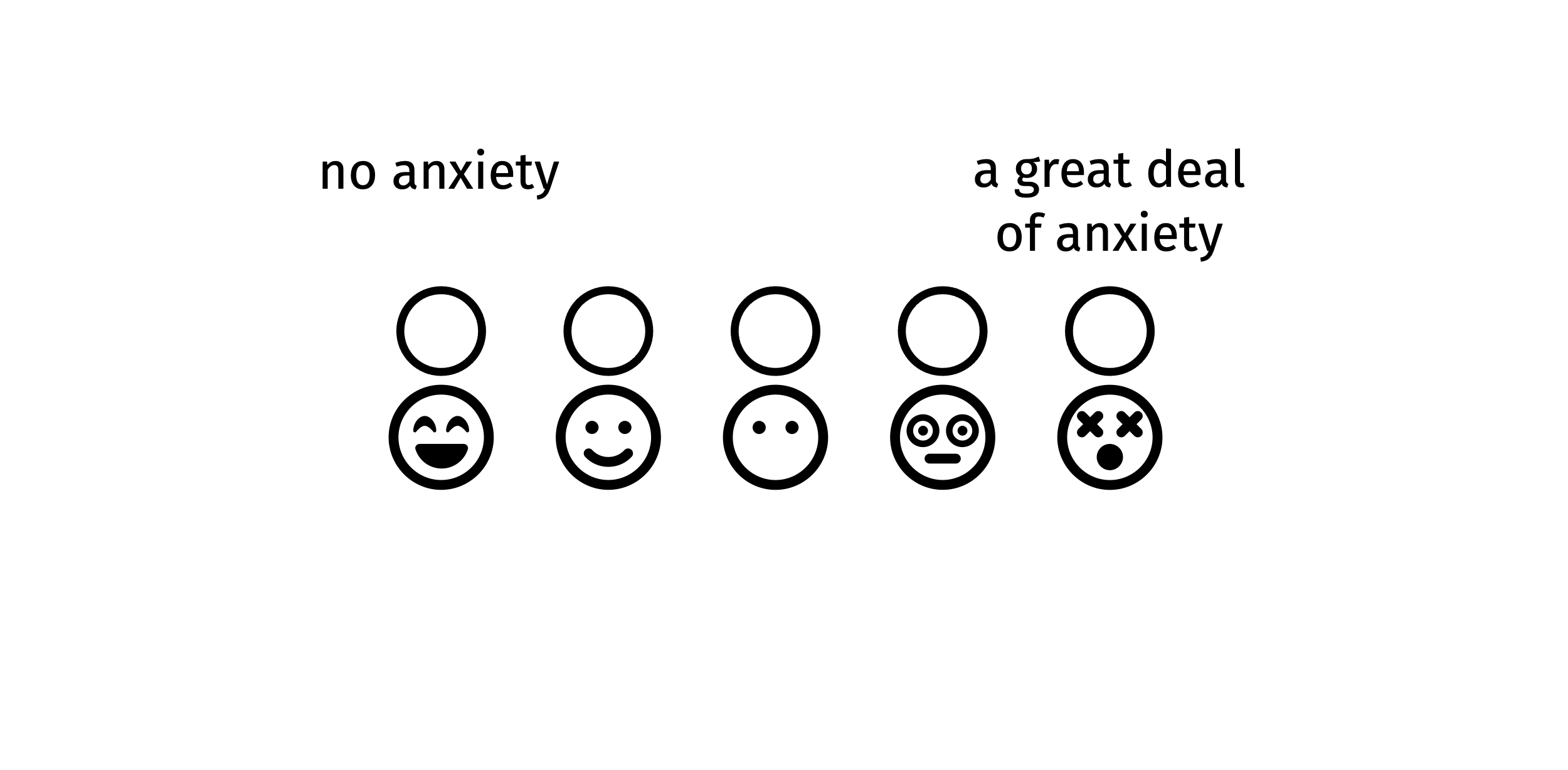


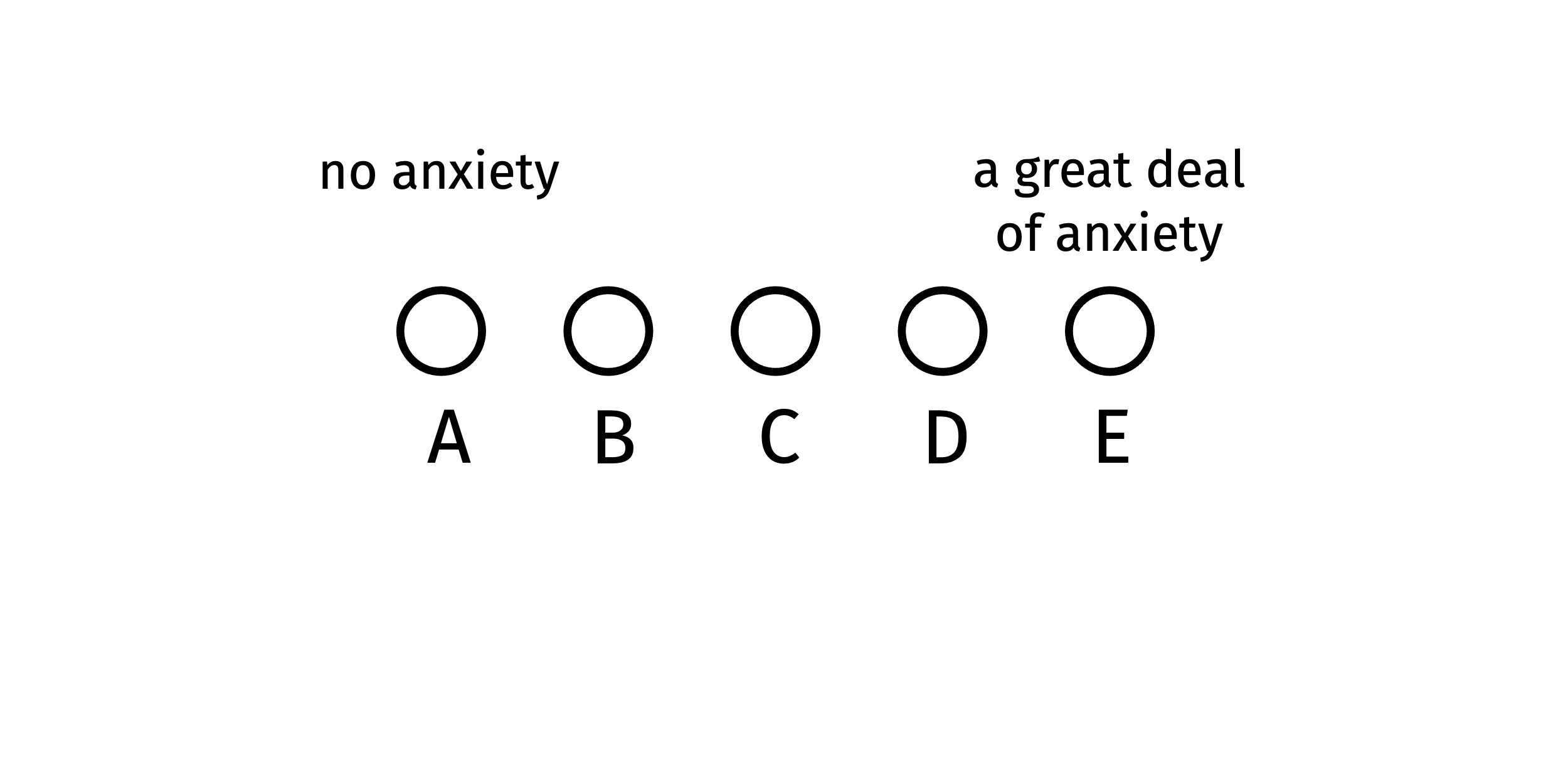
Taking the means of discrete ratings is very common—but a little strange!

# Why Likert scale ratings aren’t continuous numeric



Numbers on a Likert scale are just labels.

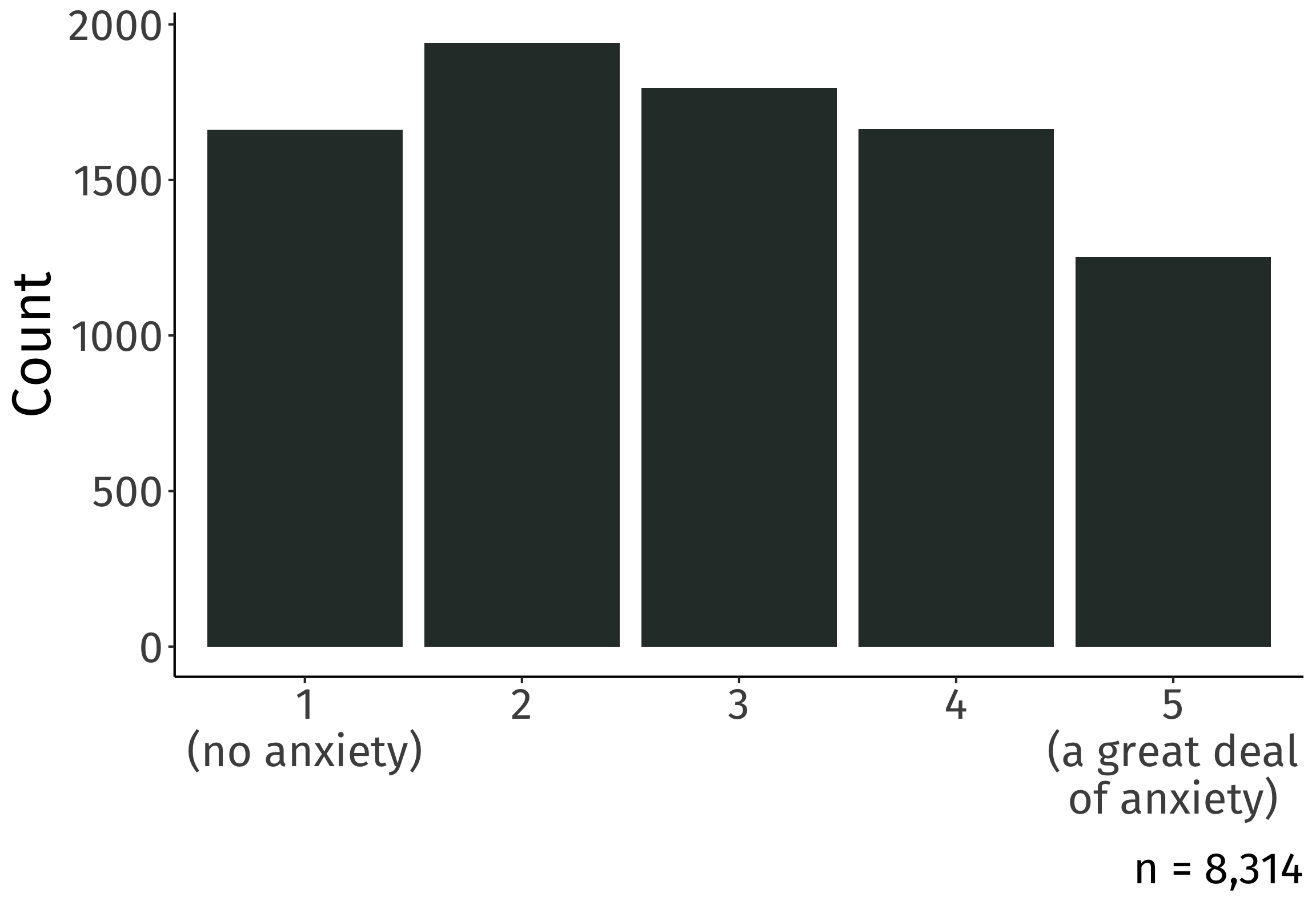




# The mistake and how you’ll avoid it

| **The mistake** | **How you’ll avoid it** |
| --- | --- |
| **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric. |  |
| **A foundational stats mistake:** Modelling categorical, ordinal data as if it were numeric. |  |
| **A foundational stats mistake:** Interpreting a significant *p*-value as evidence that an effect exists in the real world. |  |

# The data: Students’ anxiety ratings for “Going to ask my statistics teacher for individual help with material I am having difficulty understanding”.



slice(anx, 45:50)

## # A tibble: 6 × 3  
## unique\_id gender rating  
## <chr> <chr> <dbl>  
## 1 7d28c303 Female/Woman 4  
## 2 7d55383a Another Gender 4  
## 3 8116550a Female/Woman 1  
## 4 83491ff9 Female/Woman 4  
## 5 8450f8ad Male/Man 2  
## 6 876547d6 Female/Woman 3

rating looks like numbers, and R treats it like numbers, as dbl.

So it’s tempting to manipulate it like numbers.

mean(anx$rating)

## [1] 2.868054

# Remember: We are smarter than R is

Store categorical variables as factors.

anx <- anx |>   
 mutate(rating = factor(rating))

Now it’s impossible to incorrectly treat them as if they’re numeric!

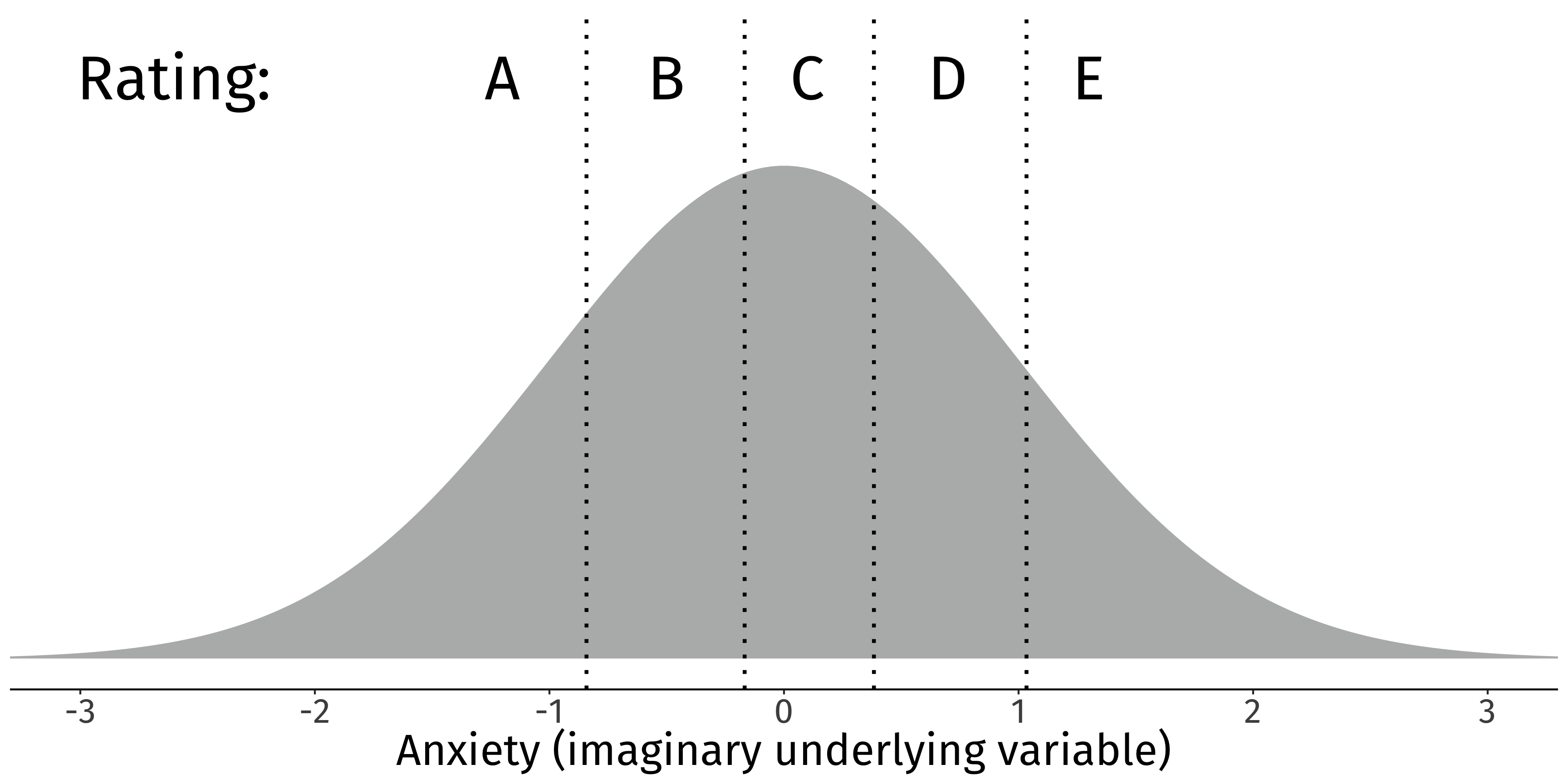
mean(anx$rating)

## [1] NA

# The mistake and how you’ll avoid it

| **The mistake** | **How you’ll avoid it** |
| --- | --- |
| **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric. | When a variable comes from a Likert scale, tell R it’s categorical using factor(). |
| **A foundational stats mistake:** Modelling categorical, ordinal data as if it were numeric. |  |
| **A foundational stats mistake:** Interpreting a significant *p*-value as evidence that an effect exists in the real world. |  |

# What ordinal regression models do



# Fit ordinal regression models with polr()

library(MASS) # MASS contains the polr() function  
  
anx\_fit1 <- polr(  
 rating ~ 1, # intercept-only model, to start  
 data = anx,   
 Hess = TRUE, method = 'probit' # ask me in the Q+A!  
)

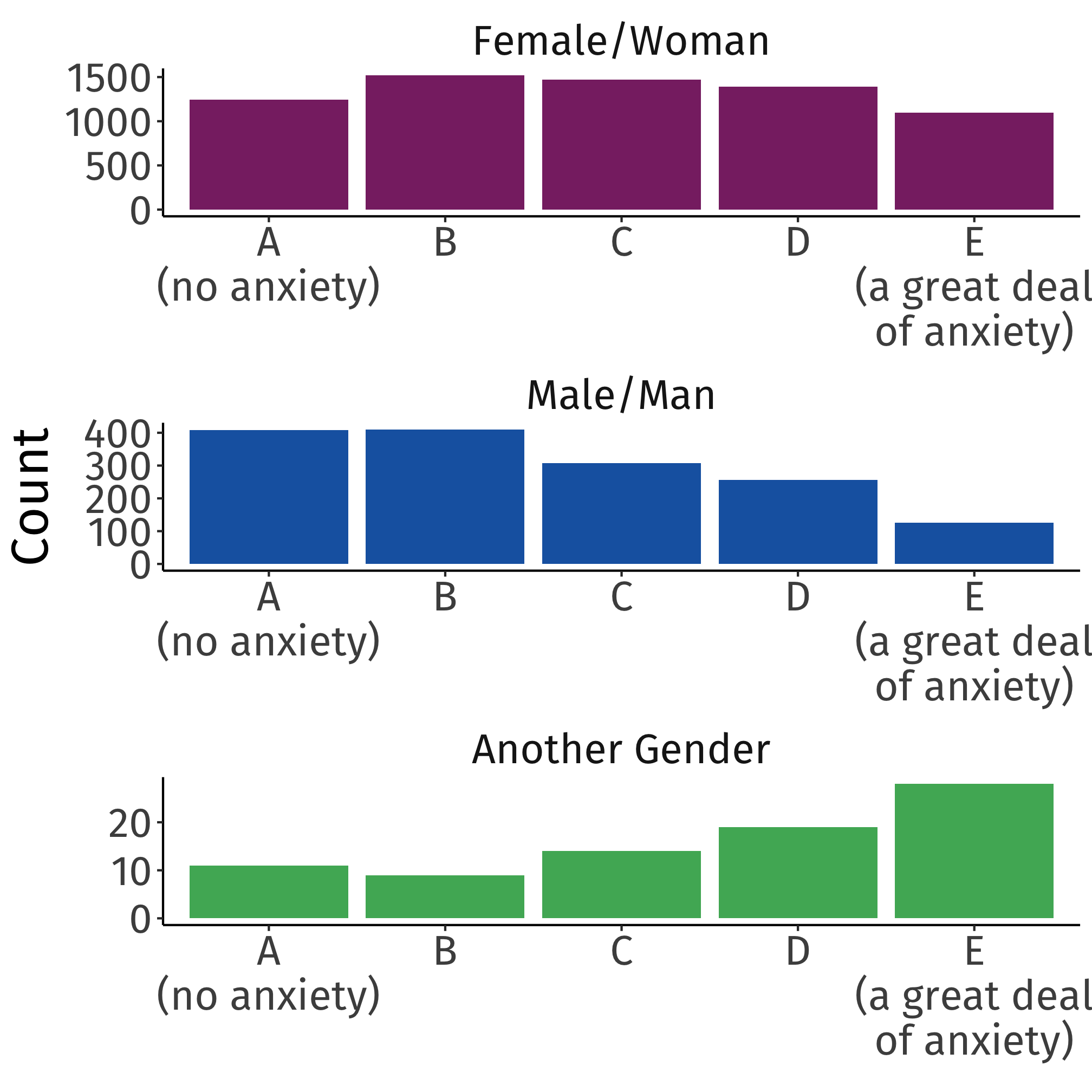
summary(anx\_fit1)

## **Intercepts:**  
## **Value** Std. Error t value   
## **A|B -0.8420** 0.0157 -53.7268  
## **B|C -0.1678** 0.0138 -12.1462  
## **C|D 0.3833** 0.0141 27.1512  
## **D|E 1.0339** 0.0168 61.6193

A grey and white graph

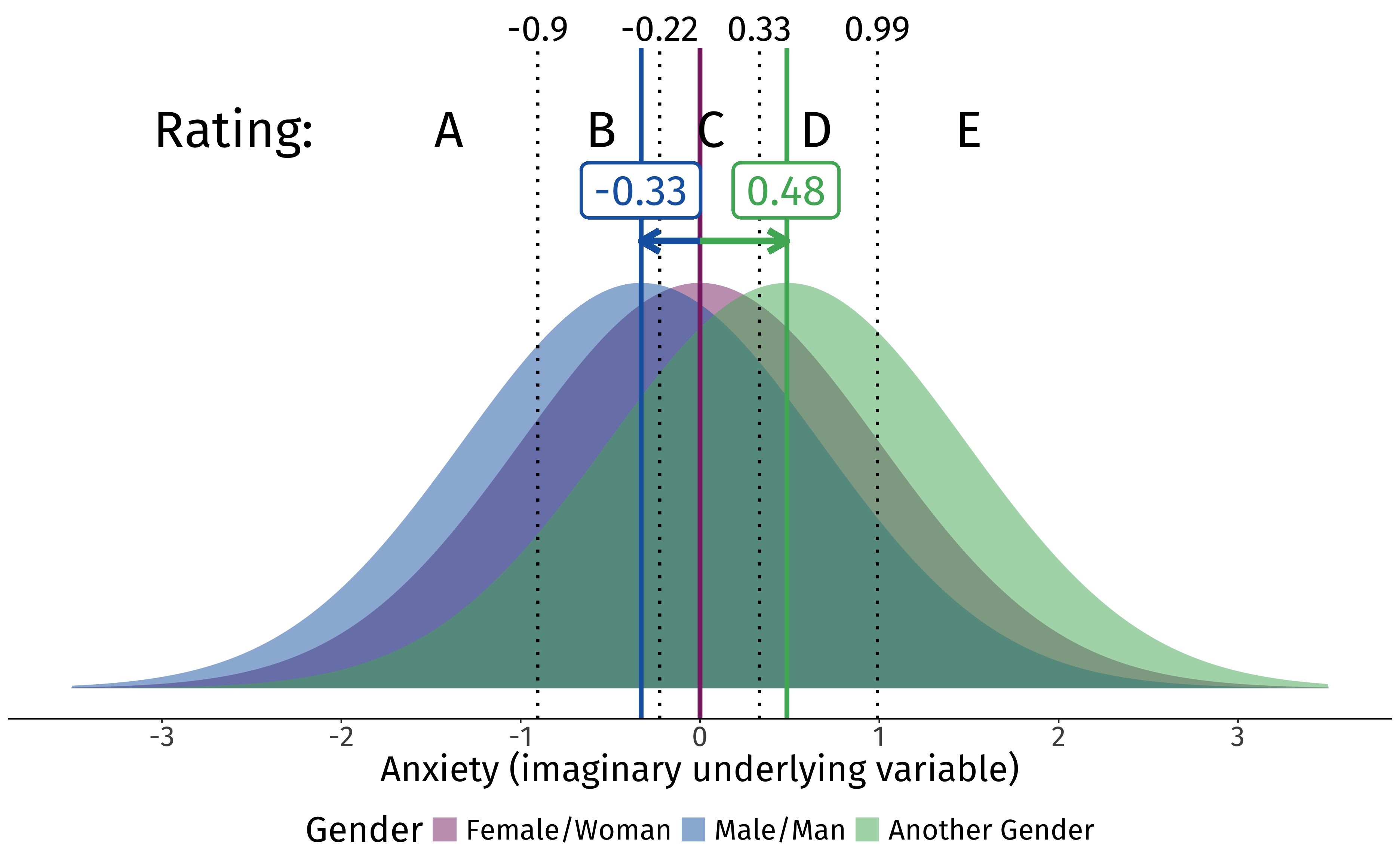
AI-generated content may be incorrect.

# How does a student’s gender affect ratings for “Going to ask my statistics teacher for individual help with material I am having difficulty understanding”?

 A graph of two people

AI-generated content may be incorrect.

**Don't flip the page until after the activity!**



# The mistake and how you’ll avoid it

| **The mistake** | **How you’ll avoid it** |
| --- | --- |
| **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric. | When a variable comes from a Likert scale, tell R it’s categorical using factor(). |
| **A foundational stats mistake:** Modelling categorical, ordinal data as if it were numeric. | Apply and interpret ordinal regression models (e.g., polr() from MASS). |
| **A foundational stats mistake:** Interpreting a significant *p*-value as evidence that an effect exists in the real world. |  |

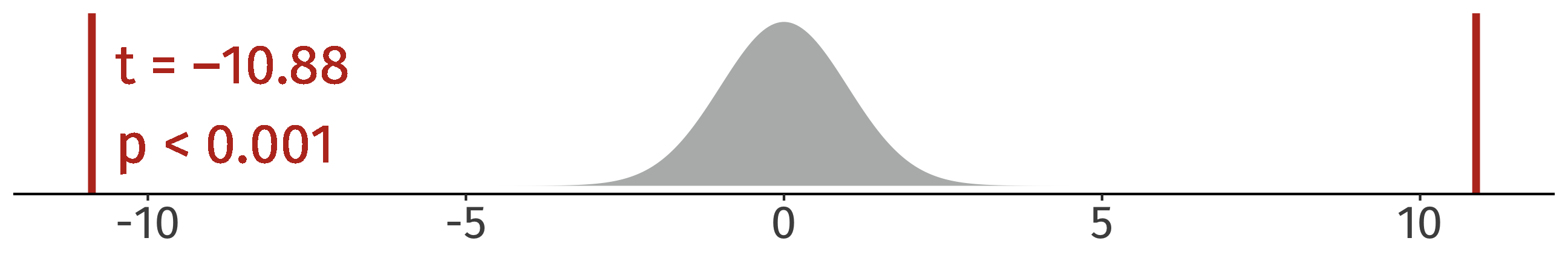
# Are the effects of gender significant?

summary(anx\_fit2)

## Coefficients:  
## Value Std. Error t value  
## genderMale/Man -0.3280 0.03015 -10.880  
## genderAnother Gender 0.4846 0.11992 4.041

No *p*-values in the model summary.

But it’s common practice to compare these *t*-values to a standard normal distribution.



A grey cone shaped graph

AI-generated content may be incorrect.

# Why don’t significant *p*-values mean an effect exists?

Because we can also get significant *p*-values when there really is *no* effect.

No difference in the true population:

A graph of different colored squares

AI-generated content may be incorrect.

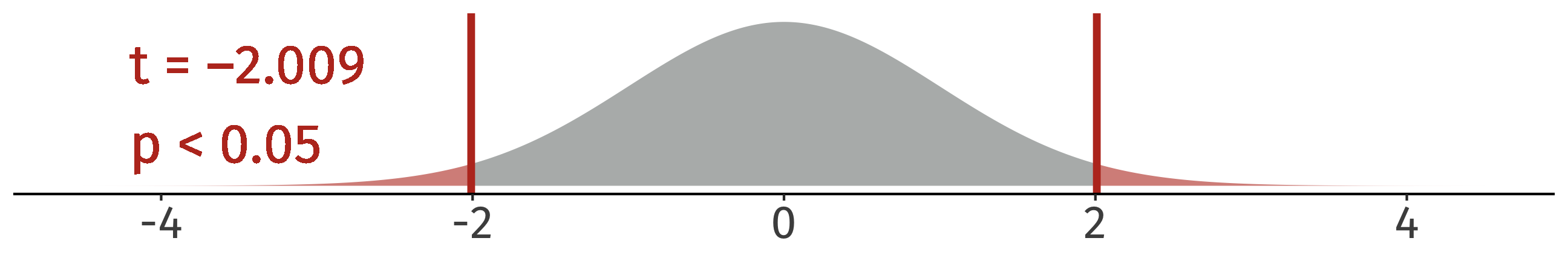
A possible random sample (*n* = 50 per group):

A graph of different colored squares

AI-generated content may be incorrect.

sim\_fit <- polr(rating ~ group, data = simdat, method = 'probit', Hess = TRUE)  
summary(sim\_fit)

## Coefficients:  
## Value Std. Error t value  
## groupGroup B -0.4479 0.2229 -2.009



So *p* is significant, but in the true population, Group A and Group B were identical!

# The mistake and how you’ll avoid it

| **The mistake** | **How you’ll avoid it** |
| --- | --- |
| **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric. | When a variable comes from a Likert scale, tell R it’s categorical using factor(). |
| **A foundational stats mistake:** Modelling categorical, ordinal data as if it were numeric. | Apply and interpret ordinal regression models (e.g., polr() from MASS). |
| **A foundational stats mistake:** Interpreting a significant *p*-value as evidence that an effect exists in the real world. | Understand that significant *p*-values can arise even if no effect exists in the real world. |

# Some really nice resources

* Jamieson’s (2004) paper [**Likert scales: How to (ab)use them.**](https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2929.2004.02012.x)
* UCLA Statistical Methods and Data Analytics’s web page [**Ordinal Logistic Regression.**](https://stats.oarc.ucla.edu/r/dae/ordinal-logistic-regression/)
* Kurz’ (2021) blog post [**Notes on the Bayesian cumulative probit.**](https://stats.oarc.ucla.edu/r/dae/ordinal-logistic-regression/)
* Vasishth and Nicenboim’s (2016) paper [**Statistical Methods for Linguistic Research: Foundational Ideas – Part I.**](https://doi.org/10.1111/lnc3.12201)
* Gelman and Hill’s (2007) book [**Data Analysis Using Regression and Multilevel/Hierarchical Models.**](https://www.cambridge.org/highereducation/books/data-analysis-using-regression-and-multilevel-hierarchical-models/32A29531C7FD730C3A68951A17C9D983)

# Plot on Slide 2 from

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2017). Distributional learning of subcategories in an artificial grammar: Category generalization and subcategory restrictions. *Journal of Memory and Language*, 97, 17–29.

# Data from

Terry, J., Ross, R. M., Nagy, T., Salgado, M., Garrido-Vásquez, P., Sarfo, J. O., Cooper, S., Buttner, A. C., Lima, T. J. S., Öztürk, İ., Akay, N., Santos, F. H., Artemenko, C., Copping, L. T., Elsherif, M. M., Milovanović, I., Cribbie, R. A., Drushlyak, M. G., Swainston, K., … Field, A. P. (2023). Data from an International Multi-Centre Study of Statistics and Mathematics Anxieties and Related Variables in University Students (the SMARVUS Dataset). *Journal of Open Psychology Data*, 11(1), 8.