Solutions Stats

Overview

N.B.: This notebook takes a relatively long time to publish with Quarto.

Some questions raised by my plots for Q10, about solutions (solutions_plots.qmd):
- Are solution scores by job category the same for all possible pairs of job groups?
- Are non-research staff significantly more likely than other groups to want a learning community?
- Are aspiring contributors significantly more likely than experienced contributors to select solutions related to learning and professional development?
- Are experienced contributors significantly more likely than aspiring contributors to select solutions related to funding?

Set seed

```
set.seed(42)
```

Import packages and utilities

```
project_root <- here::here() # requires that you be somewhere in the
# project directory (not above it)
# packages
suppressMessages(source(file.path(project_root, "scripts/packages.R")))
# functions and objects used across scripts
suppressMessages(source(file.path(project_root, "scripts/utils.R")))
```

Load data

```
solutions <- load_qualtrics_data("clean_data/solutions_Q10.tsv")
other_quant <- load_qualtrics_data("clean_data/other_quant.tsv")
```

Wrangle data

First, let's add a participant ID. We'll need to keep track of these track these since observations from the same participant are not independent. We'll need to model the participants as a random effect.

```
solutions$participantID <- seq(1, nrow(solutions))
```

Next, remove empty rows, i.e. rows from respondents who didn't receive this question. As with many questions in this survey, we can cut some corners in the code because the question was mandatory. For example, no need to worry about incomplete answers.

```
solutions_and_job <- solutions
solutions_and_job$job_category <- other_quant$job_category
names(solutions_and_job)[length(names(solutions_and_job))] <- "job_category"

nrow(solutions_and_job)
```

```
[1] 332
```

```
# from scripts/utils.R
solutions_and_job <- exclude_empty_rows(solutions_and_job, strict=TRUE)
nrow(solutions_and_job)
```

```
[1] 233
```

Good. We know by now that only 233 participants saw this question.

Here's what we have so far:

```
head(solutions_and_job)
```

```
  Computing environments       Publicity Containerization Documentation help
1            Very useful     Very useful      Very useful        Very useful
2                 Useful     Very useful      Very useful    Not very useful
3            Very useful     Very useful      Very useful        Very useful
4        Not very useful          Useful           Useful        Very useful
5                 Useful Not very useful           Useful        Very useful
7        Not very useful Not very useful      Very useful    Not very useful
  A learning community  Event planning Mentoring programs       Education
1          Very useful     Very useful        Very useful     Very useful
2               Useful  Non-applicable        Very useful     Very useful
3               Useful          Useful             Useful Not very useful
4      Not very useful          Useful    Not very useful Not very useful
5      Not very useful Not very useful             Useful     Very useful
7      Not very useful Not very useful    Not very useful Not very useful
  Legal support Industry partnerships Sustainability grants
1   Very useful           Very useful           Very useful
2   Very useful                Useful           Very useful
3   Very useful           Very useful           Very useful
4        Useful       Not very useful           Very useful
5        Useful                Useful           Very useful
7   Very useful       Not very useful           Very useful
  Help finding funding participantID         job_category
1          Very useful             1              Faculty
2               Useful             2             Post-Doc
3          Very useful             3 Other research staff
4          Very useful             4              Faculty
5               Useful             5              Faculty
7          Very useful             7              Faculty
```

Convert to long data, since this makes it easier to remove NAs and is necessary for the statistics.

```
long_data <- solutions_and_job %>%
  pivot_longer(
    cols = -c(participantID, job_category),
    names_to = "solution",
    values_to = "utility"
  )
dim(long_data)
```

```
[1] 2796    4
```

```r
head(long_data)
```

```
# A tibble: 6 x 4
  participantID job_category solution               utility
          <int> <chr>        <chr>                  <chr>
1             1 Faculty      Computing environments Very useful
2             1 Faculty      Publicity              Very useful
3             1 Faculty      Containerization       Very useful
4             1 Faculty      Documentation help     Very useful
5             1 Faculty      A learning community   Very useful
6             1 Faculty      Event planning         Very useful
```

Remove NAs.

```r
long_data <- long_data %>%
    filter(!(utility == "Non-applicable"))
dim(long_data)
```

```
[1] 2602    4
```

That removed about 200 rows, out of more than 2000. So less than 10% of the responses were "non-applicable"s.

Make utility an ordered factor. Solution and job category are not inherently ordered, but we'll make them factors, and the first factor level will be the reference level for that variable. It doesn't really matter which level we use as the reference level.

```r
long_data$utility <- factor(
  long_data$utility,
  levels = c("Not very useful", "Useful", "Very useful"),
  ordered = TRUE
)

long_data$solution <- factor(
  long_data$solution,
  levels = unique(long_data$solution)
)

long_data$job_category <- factor(
  long_data$job_category,
```

```
  levels = unique(long_data$job_category)
)

levels(long_data$solution)
```

```
 [1] "Computing environments" "Publicity"              "Containerization"
 [4] "Documentation help"     "A learning community"   "Event planning"
 [7] "Mentoring programs"     "Education"              "Legal support"
[10] "Industry partnerships"  "Sustainability grants"  "Help finding funding"
```

```
levels(long_data$job_category)
```

```
[1] "Faculty"              "Post-Doc"             "Other research staff"
[4] "Grad Student"         "Non-research Staff"   "Undergraduate"
```

Ok, so it looks like our reference levels are computing environments and faculty. That's fine. It doesn't really matter.

Create candidate models

I'd like to fit a cumulative-logit mixed model, a.k.a. an ordinal regression model, using the `clmm` function from the ordinal package. (I am not using `polr` from the MASS package because it does not allow random effects.) I know we want to include participantID as a random effect, but I'm not really sure how to model solution. I think it would be best to compare different models.

Note that the next few cells take several minutes to run.

Model 1: job_category * solution interaction

Here, I'm modeling job_category and solution as independent fixed effects, and assuming that there is also an effect from the interaction of the two. This way, we get a global slope for job_category, a global slope for solution, a global slope for the interaction (I think), and a global intercept. Adding participant as a random effect allows each participant to have their own deviation from the global intercept.

```
fit1 <- ordinal::clmm(utility ~ job_category * solution +
              (1 | participantID),
            data = long_data, link = "logit", Hess = TRUE)
```

```
Warning: (1) Hessian is numerically singular: parameters are not uniquely determined
In addition: Absolute convergence criterion was met, but relative criterion was not met
```

Hm. I get a warning that "Hessian is numerically singular: parameters are not uniquely determined" and "Absolute convergence criterion was met, but relative criterion was not met". The internet suggests that this might mean that some job-category × solution combinations have few or zero responses in one of the utility levels, so the full job_category * solution interaction is over-parameterised.

Model 2: solution as a random effect, no correlation between participant intercept and job effect

Here's another formulation. In this case, solution is another random effect, so we only get one global slope from job_category, but each solution intercept (as well as each participant intercept) is allowed to deviate from the global intercept. We assume that across solutions, the deviations in job_cateogry effect from the global effect of job_category are not correlated with that solution's intercept's deviation from the global intercept.

```r
fit2 <- ordinal::clmm(utility ~ job_category +
              (1 | solution) +
              (1 | participantID) +
              (0 + job_category | solution),
            data = long_data, link = "logit", Hess = TRUE)
```

Next, we again have 4 terms, like we did in the first model: a global intercept, slopes for job_category and solution, and a slope for the interaction. Now, we also estimate the deviance of each of these terms from the global baseline for each participant, and we also estimate the correlations between the deviations for each possible combination of the 4 terms, for each participant. Er, I think. (Helpful cheat sheet: https://stats.stackexchange.com/questions/13166/rs-lmer-cheat-sheet)

This one measures a ton of parameters… ABANDONED; NEVER CONVERGED

```r
# fit3 <- ordinal::clmm(utility ~ job_category * solution +       # fixed effects
#               (0 + job_category*solution | participantID),
#             data = long_data, link = "logit", Hess = TRUE)
```

All the models seem to be struggling a bit. Let's explore the data for a moment.

```r
# three way cross tabs (xtabs) and flatten the table
# code from: https://ladal.edu.au/tutorials/regression/regression.html
ftable(xtabs(~ job_category + solution + utility, data = long_data))
```

job_category	solution	Not very useful	Useful	Very useful
Faculty	Computing environments	12	17	29
	Publicity	19	12	24
	Containerization	19	17	18
	Documentation help	21	18	17
	A learning community	21	26	10
	Event planning	24	19	11
	Mentoring programs	24	23	8
	Education	24	21	12
	Legal support	15	28	12
	Industry partnerships	18	15	23
	Sustainability grants	3	10	44
	Help finding funding	5	13	36
Post-Doc	Computing environments	4	3	8
	Publicity	2	6	7
	Containerization	5	4	6
	Documentation help	4	6	5
	A learning community	2	9	4
	Event planning	5	3	6
	Mentoring programs	3	7	5
	Education	2	6	7
	Legal support	2	5	7
	Industry partnerships	4	3	7
	Sustainability grants	0	3	12
	Help finding funding	0	6	9
Other research staff	Computing environments	10	11	19
	Publicity	6	15	16
	Containerization	14	17	8
	Documentation help	8	14	16
	A learning community	8	19	11
	Event planning	13	14	11
	Mentoring programs	12	13	10
	Education	11	15	11
	Legal support	14	11	13
	Industry partnerships	9	12	14
	Sustainability grants	3	7	28
	Help finding funding	2	11	23
Grad Student	Computing environments	1	6	19
	Publicity	2	10	11
	Containerization	3	10	9
	Documentation help	5	8	13
	A learning community	5	9	12

	Event planning	7	6	11
	Mentoring programs	4	10	12
	Education	5	7	14
	Legal support	3	10	12
	Industry partnerships	3	11	12
	Sustainability grants	0	1	25
	Help finding funding	0	5	20
Non-research Staff	Computing environments	13	32	35
	Publicity	26	33	15
	Containerization	33	24	20
	Documentation help	19	39	26
	A learning community	11	43	31
	Event planning	29	30	16
	Mentoring programs	18	35	24
	Education	21	31	30
	Legal support	13	41	26
	Industry partnerships	23	29	18
	Sustainability grants	8	25	39
	Help finding funding	9	31	32
Undergraduate	Computing environments	0	2	5
	Publicity	0	2	4
	Containerization	1	1	4
	Documentation help	1	3	3
	A learning community	2	1	4
	Event planning	2	2	3
	Mentoring programs	0	4	3
	Education	1	4	2
	Legal support	1	3	2
	Industry partnerships	0	0	7
	Sustainability grants	0	1	5
	Help finding funding	0	2	4

Hm. Indeed, the data are sparse in places, particularly for undergraduates. Perhaps we should combine postdocs + staff researchers, as well as undergrads + grad students.

```r
combined <- long_data %>%
  mutate(
    job_category = recode(
      job_category,
      "Post-Doc" = "Postdocs and Staff Researchers",
      "Other research staff" = "Postdocs and Staff Researchers"
    )
```

```
  )

combined <- combined %>%
  mutate(
    job_category = recode(
      job_category,
      "Grad Student" = "Students",
      "Undergraduate" = "Students"
    )
  )
```

Now let's run models 1 and 2 again, but with this consolidated dataset.

Model 1b: Model 1, but with consolidated data

```
fit1b <- ordinal::clmm(utility ~ job_category * solution +
              (1 | participantID),
            data = combined, link = "logit", Hess = TRUE)
```

No warning this time, and I feel like it finished faster. My hunch is that this re-labeled dataset will lead to better results.

Model 2b: Model 2, but with consolidated data

```
fit2b <- ordinal::clmm(utility ~ job_category +
              (1 | solution) +
              (1 | participantID) +
              (0 + job_category | solution),
            data = combined, link = "logit", Hess = TRUE)
```

So, those are two fairly complex models that I think capture the important variation. Let's compare them to some simpler models.

Model 3: No job category

Let's make a null model where job category doesn't matter. (Using the consolidated data)

```r
fit3 <- ordinal::clmm(utility ~ solution +
              (1 | participantID),
            data = combined, link = "logit", Hess = TRUE)
```

Model 4: No solution category

How about a model where solution doesn't matter?

```r
fit4 <- ordinal::clmm(utility ~ job_category +
              (1 | participantID),
            data = combined, link = "logit", Hess = TRUE)
```

Model 5: job_category + solution

In this minimal model, we include job_category + solution, but without any interaction. This model says that we can predict the rating by simply adding the effect of job category and the effect of solution, with no additional effect from combining a particular job category with a particular solution.

```r
fit5 <- ordinal::clmm(utility ~ job_category + solution +
              (1 | participantID),
            data = combined, link = "logit", Hess = TRUE)
```

Model 6: no random effects

Do we really need to account for participants' individual baselines?

```r
# note clm function bc clmm is for mixed models
fit6 <- ordinal::clm(utility ~ job_category * solution,
            data = combined, link = "logit", Hess = TRUE)
```

Compare models

```
models <- list(
  "fit1"=fit1, # job_category * solution, sparser data
  "fit2"=fit2, # solution as random effect, sparser data
  "fit1b"=fit1b, # job_category * solution, denser data
  "fit2b"=fit2b, # solution as random effect, denser data
  "fit3"=fit3, # Null model: no job
  "fit4"=fit4, # Null model: no solution
  "fit5"=fit5, # Null model: no interaction
  "fit6"=fit6 # Null model: no participants
)
```

First, let's get a general sense of goodness-of-fit by looking at the AICs. You're not supposed to compare AICs for models fit to different data sets (models 1 and 2 are using the sparser data), but since I've only changed the job_category labels, but not the observations or the number of observations, I think this is ok.

```
sapply(models, function(x) round(stats::AIC(x)))
```

```
 fit1  fit2 fit1b fit2b  fit3  fit4  fit5  fit6
 4826  4847  4802  4827  4836  5094  4822  5348
```

The AICs for all the models are fairly similar, except in two cases: #4, where solution isn't doesn't matter, and job_category alone influences the response, and #6, where participant ID doesn't matter. Both of these make sense. Model 5, where job category and solution have no interaction, does fairly well. Maybe job-solution interactions are subtle.

Model 1b looks the best. According to the internet, a delta AIC of more than ten is pretty substantial, and here we have a difference of 20 between the best and second-best.

Let's check the condition number of the Hessian. I don't really understand what this is, but the clmm2 tutorial says that high numbers, say larger than say 10^4 or 10^6, indicate poor fit.

```
sapply(models, function(x)
summary(x)$info["cond.H"]
)
```

```
Warning in summary.clmm(x): Variance-covariance matrix of the parameters is not
defined
```

```
$fit1.cond.H
[1] "NaN"

$fit2.cond.H
[1] "3.9e+02"

$fit1b.cond.H
[1] "2.8e+03"

$fit2b.cond.H
[1] "2.1e+02"

$fit3.cond.H
[1] "1.5e+02"

$fit4.cond.H
[1] "1.2e+02"

$fit5.cond.H
[1] "1.6e+02"

$fit6.cond.H
[1] "3.9e+03"
```

Okay, depending on my random seed, fit1 either gives a NaN or a high value here. All the other models look decent.

Complex models vs null models

Let's use an anova to compare the two models that scored the best in terms of AIC. Since they also happen to be nested, an anova works here.

```
stats::anova(fit1b, fit5)
```

```
Likelihood ratio tests of cumulative link models:

      formula:                                                link: threshold:
fit5  utility ~ job_category + solution + (1 | participantID) logit flexible
fit1b utility ~ job_category * solution + (1 | participantID) logit flexible

      no.par    AIC  logLik LR.stat df Pr(>Chisq)
```

```
fit5      17 4822.2 -2394.1
fit1b     50 4802.0 -2351.0  86.199 33  1.221e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

That's a significant p-value. It looks like the interaction term is worth including.

Let's also double-check that participants are worth including.

```r
stats::anova(fit1b, fit6)
```

```
Likelihood ratio tests of cumulative link models:

      formula:                                             link: threshold:
fit6  utility ~ job_category * solution                    logit flexible
fit1b utility ~ job_category * solution + (1 | participantID) logit flexible

      no.par    AIC  logLik LR.stat df Pr(>Chisq)
fit6      49 5348.3 -2625.2
fit1b     50 4802.0 -2351.0  548.36  1  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Yep, definitely want to include those.

Does it matter whether we include job as a variable? Let's compare it to the model with job + solution, without an interaction term.

```r
stats::anova(fit3, fit5)
```

```
Likelihood ratio tests of cumulative link models:

     formula:                                              link: threshold:
fit3 utility ~ solution + (1 | participantID)              logit flexible
fit5 utility ~ job_category + solution + (1 | participantID) logit flexible

     no.par    AIC  logLik LR.stat df Pr(>Chisq)
fit3     14 4836.1 -2404.0
fit5     17 4822.2 -2394.1  19.902  3  0.0001779 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

It appears that job is also significant in explaining the variation in the data.

More goodness of fit evaluation

How else to evaluate the models? The `ordinal` package provides goodness-of-fit functions nominal_test and scale_test, but these only work on clm objects, not clmm objects (mixed models).

Model 2b had a similar AIC as model 5. While I can't compare model 1b and model 2b with anova, since they're not nested, I can at least glance at the standard errors of the coefficients, which I think gives me a sense of the precision of the coefficient estimates.

```
summary(fit1b$coefficients)
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-2.15349 -0.68400 -0.05181 -0.07111  0.77637  1.73451
```

```
summary(fit2b$coefficients)
```

```
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-1.38345  0.06018  0.50278  0.30830  0.91531  1.44669
```

Hm. So fit1b had the lowest AIC of all the models and is significantly better at explaining the variation than the equivalent minimal model without an interaction term. However, the coefficients of fit2b have smaller SEs than those of fit1b.

How about the log likelihoods?

```
LL <- sapply(models, function(x) x$logLik)
# These are a bit hard to read so I am reordering them
LL[order(LL)]
```

```
     fit6      fit4      fit3     fit2b      fit5      fit2     fit1b      fit1
-2625.165 -2541.072 -2404.033 -2396.590 -2394.082 -2393.642 -2350.983 -2339.214
```

In this case, surprisingly, fit1 looks best. But according to the interwebs, this can happen just from having more parameters. So I think we should probably only use this to compare models that have the same number of parameters, e.g. fit3 vs. fit4.

So, I find myself in the annoying situation of having several g-o-f tests that don't perfectly agree. However, I'm leaning toward fit1b. It had the best AIC and the second-best log-likelihood. The SEs are a little concerning, but I don't think the SEs are a super reliable indicator of g-o-f anyway(?). This model consistently had pretty good g-o-f metrics, and I think it also intuitively makes the most sense.

Let's do one more test. fit6 is the equivalent model to fit1b, but with fixed effects only. Since we can do the nominal_test and scale_test on this model, let's try it and see if it sets off any red flags.

```
nominal_test(fit6)
```

```
Tests of nominal effects

formula: utility ~ job_category * solution
                     Df  logLik    AIC    LRT Pr(>Chi)
<none>                  -2625.2 5348.3
job_category          3 -2619.8 5343.7 10.629  0.01391 *
solution             11 -2613.7 5347.3 23.021  0.01755 *
job_category:solution 47 -2590.8 5373.6 68.737  0.02098 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
scale_test(fit6)
```

```
Warning: (-1) Model failed to converge with max|grad| = 0.000305507 (tol = 1e-06)
In addition: iteration limit reached

Tests of scale effects

formula: utility ~ job_category * solution
                     Df  logLik    AIC    LRT Pr(>Chi)
<none>                  -2625.2 5348.3
job_category          3 -2619.8 5343.7 10.629  0.01391 *
solution             11 -2613.7 5347.3 23.021  0.01755 *
job_category:solution
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Ouch. That's not ideal. Maybe we can proceed with caution, and follow up with a non-parametric test on whatever trends we find? https://www.fharrell.com/post/po/

Interpreting the model results

```r
summary(fit1b)
```

```
Cumulative Link Mixed Model fitted with the Laplace approximation

formula: utility ~ job_category * solution + (1 | participantID)
data:    combined

 link  threshold nobs logLik   AIC     niter        max.grad cond.H
 logit flexible  2602 -2350.98 4801.97 10197(40741) 1.22e-03 2.8e+03

Random effects:
 Groups        Name        Variance Std.Dev.
 participantID (Intercept) 2.097    1.448
Number of groups:  participantID 232

Coefficients:
                                                                       Estimate
job_categoryPostdocs and Staff Researchers                             -0.04736
job_categoryStudents                                                    1.66906
job_categoryNon-research Staff                                         -0.08350
solutionPublicity                                                      -0.66811
solutionContainerization                                               -1.06243
solutionDocumentation help                                             -1.21045
solutionA learning community                                           -1.56910
solutionEvent planning                                                 -1.83275
solutionMentoring programs                                             -1.93070
solutionEducation                                                      -1.68608
solutionLegal support                                                  -1.18188
solutionIndustry partnerships                                          -0.68400
solutionSustainability grants                                           1.73451
solutionHelp finding funding                                            1.08082
job_categoryPostdocs and Staff Researchers:solutionPublicity            0.78228
job_categoryStudents:solutionPublicity                                 -0.49909
job_categoryNon-research Staff:solutionPublicity                       -0.74669
job_categoryPostdocs and Staff Researchers:solutionContainerization    -0.03388
job_categoryStudents:solutionContainerization                          -0.64866
job_categoryNon-research Staff:solutionContainerization                -0.42332
job_categoryPostdocs and Staff Researchers:solutionDocumentation help   0.94435
job_categoryStudents:solutionDocumentation help                        -0.35912
job_categoryNon-research Staff:solutionDocumentation help               0.53196
job_categoryPostdocs and Staff Researchers:solutionA learning community  1.02260
```

```
job_categoryStudents:solutionA learning community                        -0.10889
job_categoryNon-research Staff:solutionA learning community               1.44068
job_categoryPostdocs and Staff Researchers:solutionEvent planning         0.91886
job_categoryStudents:solutionEvent planning                              -0.22759
job_categoryNon-research Staff:solutionEvent planning                     0.35351
job_categoryPostdocs and Staff Researchers:solutionMentoring programs     1.04164
job_categoryStudents:solutionMentoring programs                           0.45731
job_categoryNon-research Staff:solutionMentoring programs                 1.24157
job_categoryPostdocs and Staff Researchers:solutionEducation              1.16078
job_categoryStudents:solutionEducation                                    0.19222
job_categoryNon-research Staff:solutionEducation                          1.12431
job_categoryPostdocs and Staff Researchers:solutionLegal support          0.54219
job_categoryStudents:solutionLegal support                               -0.35340
job_categoryNon-research Staff:solutionLegal support                      0.77637
job_categoryPostdocs and Staff Researchers:solutionIndustry partnerships  0.34019
job_categoryStudents:solutionIndustry partnerships                       -0.23308
job_categoryNon-research Staff:solutionIndustry partnerships             -0.46161
job_categoryPostdocs and Staff Researchers:solutionSustainability grants -0.05181
job_categoryStudents:solutionSustainability grants                        0.28703
job_categoryNon-research Staff:solutionSustainability grants             -1.21278
job_categoryPostdocs and Staff Researchers:solutionHelp finding funding  -0.02601
job_categoryStudents:solutionHelp finding funding                        -0.76619
job_categoryNon-research Staff:solutionHelp finding funding              -1.03639
                                                                         Std. Error
job_categoryPostdocs and Staff Researchers                                  0.49392
job_categoryStudents                                                        0.61824
job_categoryNon-research Staff                                              0.44158
solutionPublicity                                                           0.40568
solutionContainerization                                                    0.40387
solutionDocumentation help                                                  0.39881
solutionA learning community                                                0.38854
solutionEvent planning                                                      0.40245
solutionMentoring programs                                                  0.39813
solutionEducation                                                           0.39749
solutionLegal support                                                       0.39142
solutionIndustry partnerships                                               0.40142
solutionSustainability grants                                               0.44904
solutionHelp finding funding                                                0.42632
job_categoryPostdocs and Staff Researchers:solutionPublicity                0.57881
job_categoryStudents:solutionPublicity                                      0.72018
job_categoryNon-research Staff:solutionPublicity                            0.52227
job_categoryPostdocs and Staff Researchers:solutionContainerization         0.57095
job_categoryStudents:solutionContainerization                               0.71859
```

```
job_categoryNon-research Staff:solutionContainerization                     0.52130
job_categoryPostdocs and Staff Researchers:solutionDocumentation help       0.57355
job_categoryStudents:solutionDocumentation help                             0.70165
job_categoryNon-research Staff:solutionDocumentation help                   0.50803
job_categoryPostdocs and Staff Researchers:solutionA learning community     0.55805
job_categoryStudents:solutionA learning community                           0.69278
job_categoryNon-research Staff:solutionA learning community                 0.49959
job_categoryPostdocs and Staff Researchers:solutionEvent planning           0.57443
job_categoryStudents:solutionEvent planning                                 0.71107
job_categoryNon-research Staff:solutionEvent planning                       0.51964
job_categoryPostdocs and Staff Researchers:solutionMentoring programs       0.57321
job_categoryStudents:solutionMentoring programs                             0.69274
job_categoryNon-research Staff:solutionMentoring programs                   0.51079
job_categoryPostdocs and Staff Researchers:solutionEducation                0.57092
job_categoryStudents:solutionEducation                                      0.69830
job_categoryNon-research Staff:solutionEducation                            0.51032
job_categoryPostdocs and Staff Researchers:solutionLegal support            0.56944
job_categoryStudents:solutionLegal support                                  0.69631
job_categoryNon-research Staff:solutionLegal support                        0.50550
job_categoryPostdocs and Staff Researchers:solutionIndustry partnerships    0.58203
job_categoryStudents:solutionIndustry partnerships                          0.71261
job_categoryNon-research Staff:solutionIndustry partnerships                0.52272
job_categoryPostdocs and Staff Researchers:solutionSustainability grants    0.63964
job_categoryStudents:solutionSustainability grants                          0.99068
job_categoryNon-research Staff:solutionSustainability grants                0.56400
job_categoryPostdocs and Staff Researchers:solutionHelp finding funding     0.60587
job_categoryStudents:solutionHelp finding funding                           0.77727
job_categoryNon-research Staff:solutionHelp finding funding                 0.54190
                                                                         z value
job_categoryPostdocs and Staff Researchers                                -0.096
job_categoryStudents                                                       2.700
job_categoryNon-research Staff                                            -0.189
solutionPublicity                                                         -1.647
solutionContainerization                                                  -2.631
solutionDocumentation help                                                -3.035
solutionA learning community                                              -4.038
solutionEvent planning                                                    -4.554
solutionMentoring programs                                                -4.849
solutionEducation                                                         -4.242
solutionLegal support                                                     -3.019
solutionIndustry partnerships                                             -1.704
solutionSustainability grants                                              3.863
solutionHelp finding funding                                               2.535
```

```
job_categoryPostdocs and Staff Researchers:solutionPublicity                 1.352
job_categoryStudents:solutionPublicity                                      -0.693
job_categoryNon-research Staff:solutionPublicity                            -1.430
job_categoryPostdocs and Staff Researchers:solutionContainerization         -0.059
job_categoryStudents:solutionContainerization                               -0.903
job_categoryNon-research Staff:solutionContainerization                     -0.812
job_categoryPostdocs and Staff Researchers:solutionDocumentation help        1.646
job_categoryStudents:solutionDocumentation help                             -0.512
job_categoryNon-research Staff:solutionDocumentation help                    1.047
job_categoryPostdocs and Staff Researchers:solutionA learning community      1.832
job_categoryStudents:solutionA learning community                           -0.157
job_categoryNon-research Staff:solutionA learning community                  2.884
job_categoryPostdocs and Staff Researchers:solutionEvent planning            1.600
job_categoryStudents:solutionEvent planning                                 -0.320
job_categoryNon-research Staff:solutionEvent planning                        0.680
job_categoryPostdocs and Staff Researchers:solutionMentoring programs        1.817
job_categoryStudents:solutionMentoring programs                              0.660
job_categoryNon-research Staff:solutionMentoring programs                    2.431
job_categoryPostdocs and Staff Researchers:solutionEducation                 2.033
job_categoryStudents:solutionEducation                                       0.275
job_categoryNon-research Staff:solutionEducation                             2.203
job_categoryPostdocs and Staff Researchers:solutionLegal support             0.952
job_categoryStudents:solutionLegal support                                  -0.508
job_categoryNon-research Staff:solutionLegal support                         1.536
job_categoryPostdocs and Staff Researchers:solutionIndustry partnerships     0.584
job_categoryStudents:solutionIndustry partnerships                          -0.327
job_categoryNon-research Staff:solutionIndustry partnerships                -0.883
job_categoryPostdocs and Staff Researchers:solutionSustainability grants    -0.081
job_categoryStudents:solutionSustainability grants                           0.290
job_categoryNon-research Staff:solutionSustainability grants                -2.150
job_categoryPostdocs and Staff Researchers:solutionHelp finding funding     -0.043
job_categoryStudents:solutionHelp finding funding                           -0.986
job_categoryNon-research Staff:solutionHelp finding funding                 -1.912
                                                                          Pr(>|z|)
job_categoryPostdocs and Staff Researchers                                0.923611
job_categoryStudents                                                      0.006941
job_categoryNon-research Staff                                            0.850023
solutionPublicity                                                         0.099586
solutionContainerization                                                  0.008524
solutionDocumentation help                                                0.002404
solutionA learning community                                              5.38e-05
solutionEvent planning                                                    5.26e-06
solutionMentoring programs                                                1.24e-06
```

```
solutionEducation                                                    2.22e-05
solutionLegal support                                                0.002532
solutionIndustry partnerships                                        0.088390
solutionSustainability grants                                        0.000112
solutionHelp finding funding                                         0.011237
job_categoryPostdocs and Staff Researchers:solutionPublicity         0.176523
job_categoryStudents:solutionPublicity                               0.488306
job_categoryNon-research Staff:solutionPublicity                     0.152797
job_categoryPostdocs and Staff Researchers:solutionContainerization  0.952681
job_categoryStudents:solutionContainerization                        0.366692
job_categoryNon-research Staff:solutionContainerization              0.416773
job_categoryPostdocs and Staff Researchers:solutionDocumentation help 0.099662
job_categoryStudents:solutionDocumentation help                      0.608771
job_categoryNon-research Staff:solutionDocumentation help            0.295052
job_categoryPostdocs and Staff Researchers:solutionA learning community 0.066882
job_categoryStudents:solutionA learning community                    0.875099
job_categoryNon-research Staff:solutionA learning community          0.003930
job_categoryPostdocs and Staff Researchers:solutionEvent planning    0.109686
job_categoryStudents:solutionEvent planning                          0.748914
job_categoryNon-research Staff:solutionEvent planning                0.496319
job_categoryPostdocs and Staff Researchers:solutionMentoring programs 0.069187
job_categoryStudents:solutionMentoring programs                      0.509161
job_categoryNon-research Staff:solutionMentoring programs            0.015071
job_categoryPostdocs and Staff Researchers:solutionEducation         0.042035
job_categoryStudents:solutionEducation                               0.783112
job_categoryNon-research Staff:solutionEducation                     0.027585
job_categoryPostdocs and Staff Researchers:solutionLegal support     0.341017
job_categoryStudents:solutionLegal support                           0.611776
job_categoryNon-research Staff:solutionLegal support                 0.124579
job_categoryPostdocs and Staff Researchers:solutionIndustry partnerships 0.558895
job_categoryStudents:solutionIndustry partnerships                   0.743610
job_categoryNon-research Staff:solutionIndustry partnerships         0.377183
job_categoryPostdocs and Staff Researchers:solutionSustainability grants 0.935446
job_categoryStudents:solutionSustainability grants                   0.772020
job_categoryNon-research Staff:solutionSustainability grants         0.031531
job_categoryPostdocs and Staff Researchers:solutionHelp finding funding 0.965754
job_categoryStudents:solutionHelp finding funding                    0.324256
job_categoryNon-research Staff:solutionHelp finding funding          0.055813

job_categoryPostdocs and Staff Researchers
job_categoryStudents                                                 **
job_categoryNon-research Staff
solutionPublicity                                                    .
```

```
solutionContainerization                                                **
solutionDocumentation help                                              **
solutionA learning community                                            ***
solutionEvent planning                                                  ***
solutionMentoring programs                                              ***
solutionEducation                                                       ***
solutionLegal support                                                   **
solutionIndustry partnerships                                           .
solutionSustainability grants                                           ***
solutionHelp finding funding                                            *
job_categoryPostdocs and Staff Researchers:solutionPublicity
job_categoryStudents:solutionPublicity
job_categoryNon-research Staff:solutionPublicity
job_categoryPostdocs and Staff Researchers:solutionContainerization
job_categoryStudents:solutionContainerization
job_categoryNon-research Staff:solutionContainerization
job_categoryPostdocs and Staff Researchers:solutionDocumentation help   .
job_categoryStudents:solutionDocumentation help
job_categoryNon-research Staff:solutionDocumentation help
job_categoryPostdocs and Staff Researchers:solutionA learning community .
job_categoryStudents:solutionA learning community
job_categoryNon-research Staff:solutionA learning community             **
job_categoryPostdocs and Staff Researchers:solutionEvent planning
job_categoryStudents:solutionEvent planning
job_categoryNon-research Staff:solutionEvent planning
job_categoryPostdocs and Staff Researchers:solutionMentoring programs   .
job_categoryStudents:solutionMentoring programs
job_categoryNon-research Staff:solutionMentoring programs               *
job_categoryPostdocs and Staff Researchers:solutionEducation            *
job_categoryStudents:solutionEducation
job_categoryNon-research Staff:solutionEducation                        *
job_categoryPostdocs and Staff Researchers:solutionLegal support
job_categoryStudents:solutionLegal support
job_categoryNon-research Staff:solutionLegal support
job_categoryPostdocs and Staff Researchers:solutionIndustry partnerships
job_categoryStudents:solutionIndustry partnerships
job_categoryNon-research Staff:solutionIndustry partnerships
job_categoryPostdocs and Staff Researchers:solutionSustainability grants
job_categoryStudents:solutionSustainability grants
job_categoryNon-research Staff:solutionSustainability grants            *
job_categoryPostdocs and Staff Researchers:solutionHelp finding funding
job_categoryStudents:solutionHelp finding funding
job_categoryNon-research Staff:solutionHelp finding funding             .
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
                      Estimate Std. Error z value
Not very useful|Useful  -2.1535     0.3461  -6.223
Useful|Very useful       0.1717     0.3425   0.501
```

This is a lot to interpret. I'll do my best. First, let's just at the main effects, i.e. the effects of job category and solution. In the summary above, each job category is compared to Faculty, our job reference level, for the solution Computing environments, our solution reference level. The "Estimate" for job_categoryStudents is 1.66906, which indicates students have odds of e^1.67=5.3 of rating that solution at least one category higher than faculty.

The solution Publicity has a coefficient of -0.66811, indicating that faculty have odds of e^0.67=2 of rating Publicity one level lower than Computing Environments.

The interactions, e.g. job_categoryPostdocs and Staff Researchers:solutionPublicity, indicate extra log-odds only for that specific job $\times$ solution pair beyond the two main effects. So in that example, postdocs and staff researchers have an extra log-odds of 0.78228 (odds of e^0.78228=2.186) of giving publicity a higher rating than computing environments, as compared to faculty.

Interestingly, none of our p-values are super significant for interactions, meaning none of the interactions are really significant on their own. The most significant effects (three asterisks) were all solutions: A learning community (-), Event planning (-), Mentoring programs (-), Education (-), Sustainability grants (+).

So, faculty had significantly higher odds of selecting sustainability grants than computing environments; significantly lower odds of selections education, mentoring, etc. than computing environments.

One job category did get two asterisks:
Coefficients:
Estimate Std. Error z value Pr(>|z|)
job_categoryStudents 1.66906 0.61824 2.700 0.006941 **

So, students had somewhat significantly higher odds of selecting computing environments than faculty.

So, painting this with a really broad brush, we might say that responses vary across solutions more than they vary across job categories, at least in the sense that there are more significant differences within faculty than between faculty vs. students.

Since coefficients are hard to interpret, let's get contrasts using the emmeans package. The contrast essentially indicates the difference between two factors' effect sizes. So instead of comparing the coefficients by eye, we can just calculate contrasts that tell us how big the difference is, for each pair of coefficients.

Estimated marginal means

So, here's my attempt to make sense of a complicated post-hoc exploration of a complicated model. Ordinal regression with the `ordinal` package–and ordinal regression in general, I think–assumes that there is a continuous random variable–a "latent" variable–underlying the categorical outcomes. The category boundaries are then thresholds on the continuous function. The emmeans package gets estimated marginal means from your model: mean outcomes for certain variables while holding other variables constant. The emmeans function can be run in various modes that will change the reported means from the default "latent" scale (whose bounds are arbitrary) to something else. mode = "prob" will report descriptive statistics on the probability distribution of each rating. mode = "mean.class" will report the means of these distributions as probabilities on a scale of 1 to n, where n is the number of outcome categories in your data set. So if you have three outcomes, e.g. not very useful, useful, very useful, and you obtain an average rating of 2.1 for a particular solution with mode="mean.class", this means that the (estimated) average rating for that solution was 2.1, or, a teensy bit above "useful".

I'm using mode="mean.class" because I find it much easier to interpret an average rating (the sum of the probabilities of each of the three rating categories) than values on the arbitrary latent scale.

N.B.: A warning to keep in mind when using mode="prob", and I assume it also applies to mode="mean.class": https://stats.stackexchange.com/questions/615711/why-are-emmip-response-y-axis-numbers-not-probabilities-for-ordinal-regressi#:~:text=There%20are%20several%20ways%20to, I think we will be okay as long as we include job in the estimate formula?

emmeans also gives you the option to weight the means by averaging over a factor. This handy command lets us see the weights in our model. https://stats.stackexchange.com/questions/610912/emmeans-weights-for-unbalanced-groups-factors

```r
ref_grid(fit1b)@grid
```

```
                     job_category               solution .wgt.
1                         Faculty Computing environments    58
2  Postdocs and Staff Researchers Computing environments    55
3                        Students Computing environments    33
4              Non-research Staff Computing environments    80
```

5	Faculty	Publicity	55
6	Postdocs and Staff Researchers	Publicity	52
7	Students	Publicity	29
8	Non-research Staff	Publicity	74
9	Faculty	Containerization	54
10	Postdocs and Staff Researchers	Containerization	54
11	Students	Containerization	28
12	Non-research Staff	Containerization	77
13	Faculty	Documentation help	56
14	Postdocs and Staff Researchers	Documentation help	53
15	Students	Documentation help	33
16	Non-research Staff	Documentation help	84
17	Faculty	A learning community	57
18	Postdocs and Staff Researchers	A learning community	53
19	Students	A learning community	33
20	Non-research Staff	A learning community	85
21	Faculty	Event planning	54
22	Postdocs and Staff Researchers	Event planning	52
23	Students	Event planning	31
24	Non-research Staff	Event planning	75
25	Faculty	Mentoring programs	55
26	Postdocs and Staff Researchers	Mentoring programs	50
27	Students	Mentoring programs	33
28	Non-research Staff	Mentoring programs	77
29	Faculty	Education	57
30	Postdocs and Staff Researchers	Education	52
31	Students	Education	33
32	Non-research Staff	Education	82
33	Faculty	Legal support	55
34	Postdocs and Staff Researchers	Legal support	52
35	Students	Legal support	31
36	Non-research Staff	Legal support	80
37	Faculty	Industry partnerships	56
38	Postdocs and Staff Researchers	Industry partnerships	49
39	Students	Industry partnerships	33
40	Non-research Staff	Industry partnerships	70
41	Faculty	Sustainability grants	57
42	Postdocs and Staff Researchers	Sustainability grants	53
43	Students	Sustainability grants	32
44	Non-research Staff	Sustainability grants	72
45	Faculty	Help finding funding	54
46	Postdocs and Staff Researchers	Help finding funding	51
47	Students	Help finding funding	31

```
48             Non-research Staff   Help finding funding    72
```

It appears that non-research staff are weighted more heavily, and students less so, presumably because there are a lot of observations for that group and not many for the other, respectively.

```
sapply(
  c(
    "Students",
    "Non-research Staff",
    "Postdocs and Staff Researchers",
    "Faculty"
  ),
  function(x) {
    nrow(subset(combined, job_category == x))
  }
)
```

```
                      Students             Non-research Staff
                           380                            928
Postdocs and Staff Researchers                        Faculty
                           626                            668
```

First, let's explore the outcomes with different weighting schemes. I'm not cherry picking here, I'm just trying to understand the options. Let's calculate estimated marginal means for each solution, while holding job category constant. These will be really rough estimates, since we're averaging all the job categories, either equally or in proportion to their sample sizes.

(Here's a somewhat helpful explanation of weights in emmeans: https://stackoverflow.com/questions/66748520/what-is-the-difference-between-weights-cell-and-weights-proportional-in-r-pa)

```
# code copied from https://cran.r-project.org/web/packages/emmeans/vignettes/messy-data.html#
sapply(c("equal", "prop", "outer", "cells", "flat"), \(w)
    emmeans(fit1b, ~ solution, weights = w) |> predict()) |> head()
```

```
NOTE: Results may be misleading due to involvement in interactions
NOTE: Results may be misleading due to involvement in interactions
NOTE: Results may be misleading due to involvement in interactions
NOTE: Results may be misleading due to involvement in interactions
NOTE: Results may be misleading due to involvement in interactions
```

```
           equal       prop      outer      cells        flat
[1,]  1.37546661  1.1934942  1.1934942  1.1935455  1.37546661
[2,]  0.59148638  0.3743983  0.3743983  0.3738146  0.59148638
[3,]  0.03657354 -0.1227918 -0.1227918 -0.1411864  0.03657354
[4,]  0.44431303  0.3475157  0.3475157  0.3487806  0.44431303
[5,]  0.39495919  0.3683255  0.3683255  0.3802930  0.39495919
[6,] -0.19609423 -0.3253567 -0.3253567 -0.3217714 -0.19609423
```

We only get two sets of estimates: equal/flat gives us the estimates where all means are given equal weight. Prop, outer, and cells give us another set of estimates, where each prediction is given the weights proportional to sample size. At least, I think that's how it works.

Let's look at the average ratings by job category (our weighting scheme here doesn't matter, because we're splitting it up by job, not averaging over job).

```r
all_means <- emmeans(fit1b, ~ solution | job_category, mode="mean.class")
summary(all_means) %>%
 arrange(desc(mean.class))
```

```
job_category = Faculty:
 solution                mean.class     SE  df asymp.LCL asymp.UCL
 Sustainability grants         2.81 0.0648 Inf      2.68      2.93
 Help finding funding          2.67 0.0898 Inf      2.50      2.85
 Computing environments        2.35 0.1170 Inf      2.12      2.58
 Publicity                     2.12 0.1250 Inf      1.87      2.36
 Industry partnerships         2.11 0.1240 Inf      1.87      2.35
 Containerization              1.97 0.1250 Inf      1.73      2.22
 Legal support                 1.93 0.1190 Inf      1.70      2.16
 Documentation help            1.92 0.1220 Inf      1.68      2.16
 A learning community          1.79 0.1160 Inf      1.56      2.02
 Education                     1.75 0.1190 Inf      1.52      1.98
 Event planning                1.70 0.1190 Inf      1.47      1.93
 Mentoring programs            1.66 0.1150 Inf      1.44      1.89

job_category = Postdocs and Staff Researchers:
 solution                mean.class     SE  df asymp.LCL asymp.UCL
 Sustainability grants         2.79 0.0699 Inf      2.65      2.93
 Help finding funding          2.66 0.0929 Inf      2.47      2.84
 Publicity                     2.38 0.1180 Inf      2.14      2.61
 Computing environments        2.34 0.1230 Inf      2.10      2.58
 Documentation help            2.24 0.1230 Inf      2.00      2.49
 Industry partnerships         2.22 0.1280 Inf      1.97      2.47
```

```
 Education                     2.15 0.1240 Inf      1.91      2.40
 A learning community          2.14 0.1210 Inf      1.91      2.38
 Legal support                 2.11 0.1260 Inf      1.86      2.36
 Mentoring programs            2.02 0.1270 Inf      1.77      2.27
 Event planning                2.01 0.1250 Inf      1.76      2.26
 Containerization              1.94 0.1220 Inf      1.70      2.18

job_category = Students:
 solution                mean.class     SE  df asymp.LCL asymp.UCL
 Sustainability grants         2.97 0.0250 Inf      2.92      3.02
 Help finding funding          2.84 0.0743 Inf      2.70      2.99
 Computing environments        2.80 0.0877 Inf      2.62      2.97
 Industry partnerships         2.59 0.1310 Inf      2.33      2.85
 Publicity                     2.52 0.1440 Inf      2.23      2.80
 Mentoring programs            2.42 0.1450 Inf      2.13      2.70
 Education                     2.41 0.1490 Inf      2.12      2.70
 Legal support                 2.40 0.1500 Inf      2.10      2.69
 Documentation help            2.39 0.1510 Inf      2.09      2.68
 A learning community          2.35 0.1530 Inf      2.05      2.65
 Containerization              2.34 0.1610 Inf      2.02      2.66
 Event planning                2.22 0.1640 Inf      1.89      2.54

job_category = Non-research Staff:
 solution                mean.class     SE  df asymp.LCL asymp.UCL
 Sustainability grants         2.50 0.0922 Inf      2.32      2.68
 Help finding funding          2.34 0.0992 Inf      2.15      2.53
 Computing environments        2.32 0.0965 Inf      2.14      2.51
 A learning community          2.28 0.0930 Inf      2.10      2.46
 Legal support                 2.18 0.0973 Inf      1.99      2.37
 Education                     2.13 0.0986 Inf      1.93      2.32
 Documentation help            2.08 0.0964 Inf      1.89      2.27
 Mentoring programs            2.08 0.0991 Inf      1.88      2.27
 Industry partnerships         1.91 0.1050 Inf      1.71      2.12
 Publicity                     1.82 0.1010 Inf      1.62      2.01
 Event planning                1.79 0.1010 Inf      1.60      1.99
 Containerization              1.79 0.1010 Inf      1.59      1.99

Confidence level used: 0.95
```

Here, we see that "a learning community" is more popular among non-research staff than among other groups. So, we expect that if all groups are weighted equally, "a learning community" will be less popular than if we weight the means by sample size.

Hmm. When I try running the command below, to get global solution ratings averaged over job, the command fails when I include mode="mean.class". It says no weighting information is given.

```
summary(emmeans(fit1b, ~ solution, weights = "equal", mode = "mean.class")) %>%
  arrange(desc(mean.class))
```

```
Warning in emmeans(fit1b, ~solution, weights = "equal", mode = "mean.class"):
'weights' requested but no weighting information is available
```

```
NOTE: Results may be misleading due to involvement in interactions
```

```
 solution                mean.class     SE  df asymp.LCL asymp.UCL
 Sustainability grants         2.77 0.0338 Inf      2.70      2.83
 Help finding funding          2.63 0.0449 Inf      2.54      2.72
 Computing environments        2.45 0.0535 Inf      2.35      2.56
 Industry partnerships         2.21 0.0611 Inf      2.09      2.33
 Publicity                     2.21 0.0616 Inf      2.09      2.33
 Documentation help            2.16 0.0625 Inf      2.04      2.28
 Legal support                 2.16 0.0624 Inf      2.03      2.28
 A learning community          2.14 0.0613 Inf      2.02      2.26
 Education                     2.11 0.0619 Inf      1.99      2.23
 Mentoring programs            2.05 0.0613 Inf      1.93      2.17
 Containerization              2.01 0.0647 Inf      1.89      2.14
 Event planning                1.93 0.0647 Inf      1.80      2.06

Results are averaged over the levels of: job_category
Confidence level used: 0.95
```

I think we can achieve the same result by running emmeans() on an emmGrid object, which doesn't do any predictions, it just collapses/averages the means already obtained in the first emmeans call. Let's check this without the mean.class argument.

```
t <- emmeans(fit1b, ~ solution | job_category)
t2 <- emmeans(t, ~ solution)
```

```
NOTE: Results may be misleading due to involvement in interactions
```

```r
t3 <- emmeans(fit1b, ~ solution)
```

```
NOTE: Results may be misleading due to involvement in interactions
```

```r
head(summary(t2))
```

```
 solution                 emmean    SE  df asymp.LCL asymp.UCL
 Computing environments  1.3755 0.193 Inf    0.9968     1.754
 Publicity               0.5915 0.185 Inf    0.2288     0.954
 Containerization        0.0366 0.183 Inf   -0.3214     0.395
 Documentation help      0.4443 0.179 Inf    0.0940     0.795
 A learning community    0.3950 0.175 Inf    0.0523     0.738
 Event planning         -0.1961 0.182 Inf   -0.5523     0.160

Results are averaged over the levels of: job_category
Confidence level used: 0.95
```

```r
head(summary(t3))
```

```
 solution                 emmean    SE  df asymp.LCL asymp.UCL
 Computing environments  1.3755 0.193 Inf    0.9968     1.754
 Publicity               0.5915 0.185 Inf    0.2288     0.954
 Containerization        0.0366 0.183 Inf   -0.3214     0.395
 Documentation help      0.4443 0.179 Inf    0.0940     0.795
 A learning community    0.3950 0.175 Inf    0.0523     0.738
 Event planning         -0.1961 0.182 Inf   -0.5523     0.160

Results are averaged over the levels of: job_category
Confidence level used: 0.95
```

The results are the same. So I think we can use this method to get weighted global means on the mean.class scale. Let's try it.

```r
overall_equal <- emmeans(all_means, ~ solution, weights = "equal")
```

```
Warning in emmeans(all_means, ~solution, weights = "equal"): 'weights'
requested but no weighting information is available
```

```
NOTE: Results may be misleading due to involvement in interactions
```

Still getting the error. Here are some hail mary attempts to troubleshoot.

```r
head(summary(emmeans(all_means, ~ solution, by = NULL)))
```

```
NOTE: Results may be misleading due to involvement in interactions
```

```
 solution                mean.class     SE  df asymp.LCL asymp.UCL
 Computing environments        2.45 0.0535 Inf      2.35      2.56
 Publicity                     2.21 0.0616 Inf      2.09      2.33
 Containerization              2.01 0.0647 Inf      1.89      2.14
 Documentation help            2.16 0.0625 Inf      2.04      2.28
 A learning community          2.14 0.0613 Inf      2.02      2.26
 Event planning                1.93 0.0647 Inf      1.80      2.06

Results are averaged over the levels of: job_category
Confidence level used: 0.95
```

```r
head(summary(emmeans(all_means, ~ solution, weights = NULL)))
```

```
NOTE: Results may be misleading due to involvement in interactions
```

```
 solution                mean.class     SE  df asymp.LCL asymp.UCL
 Computing environments        2.45 0.0535 Inf      2.35      2.56
 Publicity                     2.21 0.0616 Inf      2.09      2.33
 Containerization              2.01 0.0647 Inf      1.89      2.14
 Documentation help            2.16 0.0625 Inf      2.04      2.28
 A learning community          2.14 0.0613 Inf      2.02      2.26
 Event planning                1.93 0.0647 Inf      1.80      2.06

Results are averaged over the levels of: job_category
Confidence level used: 0.95
```

```r
head(summary(emmeans(all_means, ~ solution, weights = "prop")))
```

```
Warning in emmeans(all_means, ~solution, weights = "prop"): 'weights' requested
but no weighting information is available
```

```
NOTE: Results may be misleading due to involvement in interactions
```

```
 solution                mean.class     SE  df asymp.LCL asymp.UCL
 Computing environments        2.45 0.0535 Inf      2.35      2.56
 Publicity                     2.21 0.0616 Inf      2.09      2.33
 Containerization              2.01 0.0647 Inf      1.89      2.14
 Documentation help            2.16 0.0625 Inf      2.04      2.28
 A learning community          2.14 0.0613 Inf      2.02      2.26
 Event planning                1.93 0.0647 Inf      1.80      2.06

Results are averaged over the levels of: job_category
Confidence level used: 0.95
```

```r
head(summary(emmeans(all_means, ~ solution, weights = ref_grid(fit1b)@grid)))
```

```
NOTE: Results may be misleading due to involvement in interactions
```

```
 solution                mean.class     SE  df asymp.LCL asymp.UCL
 Computing environments        2.45 0.0535 Inf      2.35      2.56
 Publicity                     2.21 0.0616 Inf      2.09      2.33
 Containerization              2.01 0.0647 Inf      1.89      2.14
 Documentation help            2.16 0.0625 Inf      2.04      2.28
 A learning community          2.14 0.0613 Inf      2.02      2.26
 Event planning                1.93 0.0647 Inf      1.80      2.06

Results are averaged over the levels of: job_category
Confidence level used: 0.95
```

```r
head(summary(emmeans(all_means, ~ solution, weights = "show.levels")))
```

```
Warning in emmeans(all_means, ~solution, weights = "show.levels"): 'weights'
requested but no weighting information is available
```

```
NOTE: Results may be misleading due to involvement in interactions
```

```
 solution                mean.class     SE  df asymp.LCL asymp.UCL
 Computing environments        2.45 0.0535 Inf      2.35      2.56
 Publicity                     2.21 0.0616 Inf      2.09      2.33
 Containerization              2.01 0.0647 Inf      1.89      2.14
 Documentation help            2.16 0.0625 Inf      2.04      2.28
 A learning community          2.14 0.0613 Inf      2.02      2.26
 Event planning                1.93 0.0647 Inf      1.80      2.06
```

```
Results are averaged over the levels of: job_category
Confidence level used: 0.95
```

Okay, so I'm having a really hard time getting emmeans to incorporate weights when averaging over the estimates made in mean.class mode. I guess I will just use the default log scale. Maybe it's just not possible to incorporate weights when using the mean.class scale, because of math that I don't understand.

```
overall_equal <- emmeans(fit1b, ~ solution, weights = "equal")
```

```
NOTE: Results may be misleading due to involvement in interactions
```

```
summary(overall_equal) %>%
  arrange(desc(emmean))
```

```
 solution                 emmean    SE  df asymp.LCL asymp.UCL
 Sustainability grants    2.8656 0.262 Inf    2.3524     3.379
 Help finding funding     1.9991 0.206 Inf    1.5959     2.402
 Computing environments   1.3755 0.193 Inf    0.9968     1.754
 Industry partnerships    0.6028 0.185 Inf    0.2394     0.966
 Publicity                0.5915 0.185 Inf    0.2288     0.954
 Documentation help       0.4443 0.179 Inf    0.0940     0.795
 Legal support            0.4349 0.178 Inf    0.0858     0.784
 A learning community     0.3950 0.175 Inf    0.0523     0.738
 Education                0.3087 0.178 Inf   -0.0408     0.658
 Mentoring programs       0.1299 0.177 Inf   -0.2176     0.477
 Containerization         0.0366 0.183 Inf   -0.3214     0.395
 Event planning          -0.1961 0.182 Inf   -0.5523     0.160

Results are averaged over the levels of: job_category
Confidence level used: 0.95
```

```
overall_prop <- emmeans(fit1b, ~ solution, weights = "prop")
```

```
NOTE: Results may be misleading due to involvement in interactions
```

```
summary(overall_prop) %>%
  arrange(desc(emmean))
```

```
 solution                emmean    SE  df asymp.LCL asymp.UCL
 Sustainability grants    2.525 0.216 Inf    2.1017    2.9481
 Help finding funding     1.787 0.187 Inf    1.4195    2.1536
 Computing environments   1.193 0.176 Inf    0.8476    1.5394
 Industry partnerships    0.393 0.175 Inf    0.0501    0.7352
 Publicity                0.374 0.173 Inf    0.0356    0.7132
 A learning community     0.368 0.163 Inf    0.0479    0.6888
 Legal support            0.367 0.167 Inf    0.0396    0.6951
 Documentation help       0.348 0.167 Inf    0.0199    0.6751
 Education                0.216 0.168 Inf   -0.1131    0.5446
 Mentoring programs       0.023 0.168 Inf   -0.3057    0.3517
 Containerization        -0.123 0.171 Inf   -0.4586    0.2130
 Event planning          -0.325 0.171 Inf   -0.6610    0.0103

Results are averaged over the levels of: job_category
Confidence level used: 0.95
```

When we use the default weighting of "equal", "A learning community" is #8, but with "prop" weighting, it rises to #6. This makes sense, as discussed above.

The more that I think about it, the more I feel like we shouldn't even report the global emms– just emms by job. It may do more harm than good to average over a factor that we've already established is important. So let's look at emms by job.

```r
# This yields the same results: emmeans(fit1b, ~ solution | job_category, mode = "mean.class"
emm <- summary(emmeans(fit1b, ~ solution * job_category, mode = "mean.class"))
emm
```

```
 solution               job_category                    mean.class     SE  df
 Computing environments Faculty                               2.35 0.1170 Inf
 Publicity              Faculty                               2.12 0.1250 Inf
 Containerization       Faculty                               1.97 0.1250 Inf
 Documentation help     Faculty                               1.92 0.1220 Inf
 A learning community   Faculty                               1.79 0.1160 Inf
 Event planning         Faculty                               1.70 0.1190 Inf
 Mentoring programs     Faculty                               1.66 0.1150 Inf
 Education              Faculty                               1.75 0.1190 Inf
 Legal support          Faculty                               1.93 0.1190 Inf
 Industry partnerships  Faculty                               2.11 0.1240 Inf
 Sustainability grants  Faculty                               2.81 0.0648 Inf
 Help finding funding   Faculty                               2.67 0.0898 Inf
 Computing environments Postdocs and Staff Researchers       2.34 0.1230 Inf
```

```
Publicity              Postdocs and Staff Researchers       2.38 0.1180 Inf
Containerization       Postdocs and Staff Researchers       1.94 0.1220 Inf
Documentation help     Postdocs and Staff Researchers       2.24 0.1230 Inf
A learning community   Postdocs and Staff Researchers       2.14 0.1210 Inf
Event planning         Postdocs and Staff Researchers       2.01 0.1250 Inf
Mentoring programs     Postdocs and Staff Researchers       2.02 0.1270 Inf
Education              Postdocs and Staff Researchers       2.15 0.1240 Inf
Legal support          Postdocs and Staff Researchers       2.11 0.1260 Inf
Industry partnerships  Postdocs and Staff Researchers       2.22 0.1280 Inf
Sustainability grants  Postdocs and Staff Researchers       2.79 0.0699 Inf
Help finding funding   Postdocs and Staff Researchers       2.66 0.0929 Inf
Computing environments Students                             2.80 0.0877 Inf
Publicity              Students                             2.52 0.1440 Inf
Containerization       Students                             2.34 0.1610 Inf
Documentation help     Students                             2.39 0.1510 Inf
A learning community   Students                             2.35 0.1530 Inf
Event planning         Students                             2.22 0.1640 Inf
Mentoring programs     Students                             2.42 0.1450 Inf
Education              Students                             2.41 0.1490 Inf
Legal support          Students                             2.40 0.1500 Inf
Industry partnerships  Students                             2.59 0.1310 Inf
Sustainability grants  Students                             2.97 0.0250 Inf
Help finding funding   Students                             2.84 0.0743 Inf
Computing environments Non-research Staff                   2.32 0.0965 Inf
Publicity              Non-research Staff                   1.82 0.1010 Inf
Containerization       Non-research Staff                   1.79 0.1010 Inf
Documentation help     Non-research Staff                   2.08 0.0964 Inf
A learning community   Non-research Staff                   2.28 0.0930 Inf
Event planning         Non-research Staff                   1.79 0.1010 Inf
Mentoring programs     Non-research Staff                   2.08 0.0991 Inf
Education              Non-research Staff                   2.13 0.0986 Inf
Legal support          Non-research Staff                   2.18 0.0973 Inf
Industry partnerships  Non-research Staff                   1.91 0.1050 Inf
Sustainability grants  Non-research Staff                   2.50 0.0922 Inf
Help finding funding   Non-research Staff                   2.34 0.0992 Inf
asymp.LCL asymp.UCL
     2.12      2.58
     1.87      2.36
     1.73      2.22
     1.68      2.16
     1.56      2.02
     1.47      1.93
     1.44      1.89
```

```
                                   35



     1.52      1.98
     1.70      2.16
     1.87      2.35
     2.68      2.93
     2.50      2.85
     2.10      2.58
     2.14      2.61
     1.70      2.18
     2.00      2.49
     1.91      2.38
     1.76      2.26
     1.77      2.27
     1.91      2.40
     1.86      2.36
     1.97      2.47
     2.65      2.93
     2.47      2.84
     2.62      2.97
     2.23      2.80
     2.02      2.66
     2.09      2.68
     2.05      2.65
     1.89      2.54
     2.13      2.70
     2.12      2.70
     2.10      2.69
     2.33      2.85
     2.92      3.02
     2.70      2.99
     2.14      2.51
     1.62      2.01
     1.59      1.99
     1.89      2.27
     2.10      2.46
     1.60      1.99
     1.88      2.27
     1.93      2.32
     1.99      2.37
     1.71      2.12
     2.32      2.68
     2.15      2.53

Confidence level used: 0.95
```

Plot emms

Plot the results.

```
emm_clean <- emm %>%
  rename(mean = mean.class,
         lwr  = asymp.LCL,
         upr  = asymp.UCL) %>%
  mutate(across(c(mean, lwr, upr), as.numeric))

# Use a common ordering of solutions (here, overall mean w equal weighting)
solns_ordered <- summary(emmeans(fit1b, ~ solution, weights = "equal")) %>%
  arrange(emmean) %>% # don't do desc() bc these will be flipped later w coord_flip()
  pull(solution) %>%
  as.character()
```

```
NOTE: Results may be misleading due to involvement in interactions
```

```
emm_clean <- emm_clean %>%
  mutate(solution = factor(solution, levels = solns_ordered))
```

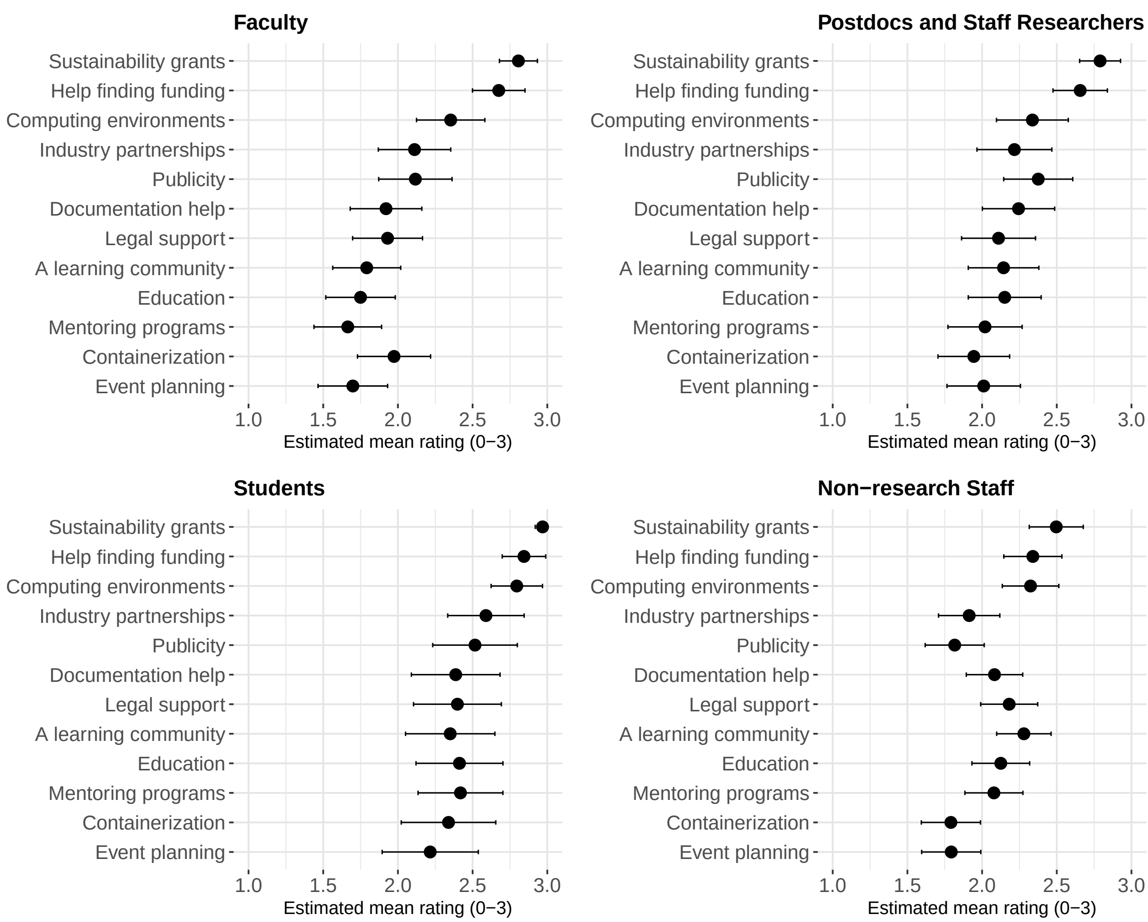
```
make_plot <- function(df, jc) {
  ggplot(filter(df, job_category == jc),
         aes(x = solution, y = mean)) +
    geom_errorbar(aes(ymin = lwr, ymax = upr),
                  width = .15, linewidth = .4) +
    geom_point(size = 3) +
    ylim(c(1, 3)) +
    labs(title = jc, x = NULL, y = "Estimated mean rating (0-3)") +
    coord_flip() +                        # solutions run down the y-axis
    theme(
      plot.title = element_text(face = "bold"),
      axis.text.x = element_text(
        size = 12
      ),
      axis.text.y = element_text(
        size = 12
        ),
      panel.background = element_blank(),
      panel.grid =
        element_line(
```

```
            linetype = "solid",
            color = "gray90"
            ),
        plot.margin = unit(c(0.3, 0.3, 0.3, 0.3), "cm")
    )
}

plots <- lapply(unique(emm_clean$job_category),
                make_plot, df = emm_clean)

composite_plot <- wrap_plots(plots, ncol = 2)

composite_plot
```



```
save_plot("solns_points1.tiff", 12, 10, p=composite_plot)
```

Let's try combining them all on one plot.

```
soln_levels <- levels(emm_clean$solution)
interleaved <- as.vector(rbind(paste0(soln_levels, "_sp"), soln_levels))
interleaved[length(interleaved)+1] <- "padding_sp"
```

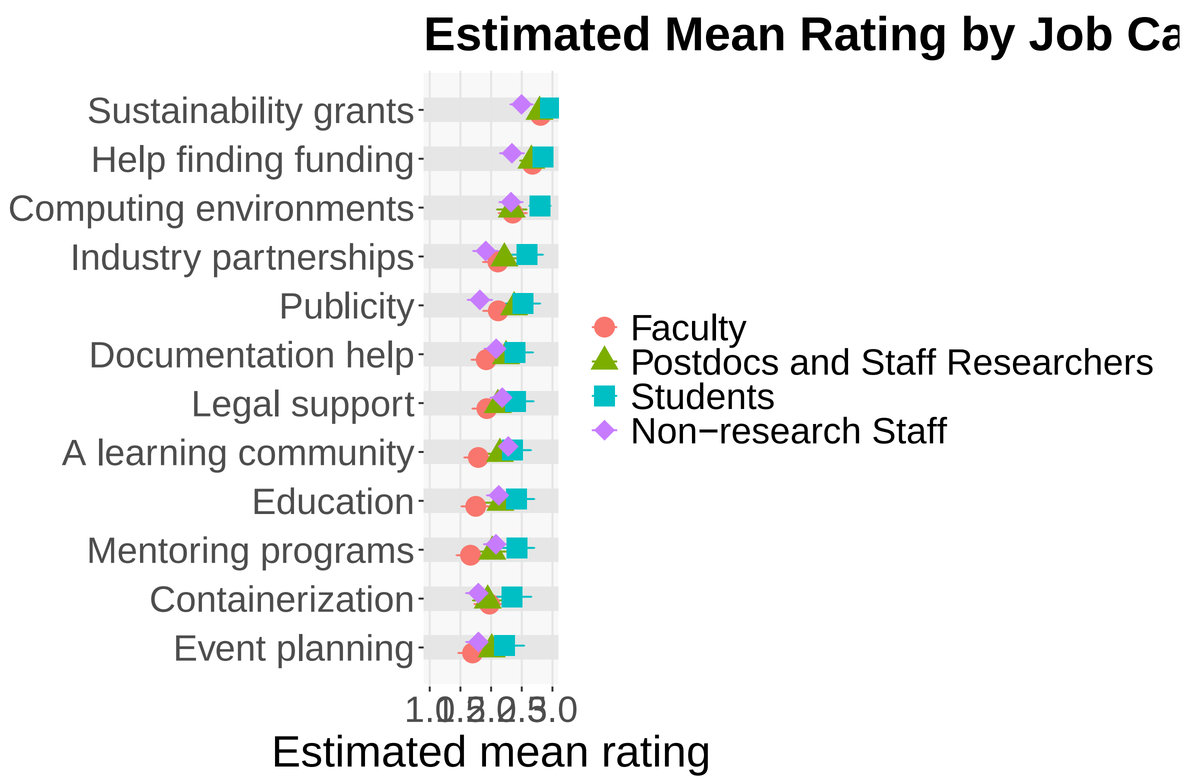
```
# Define a position dodge object to ensure points and error bars align
pd <- position_dodge(width = 0.6)

# one stripe per real category row
bg <- tibble(cat = factor(interleaved, levels = interleaved)) %>%
  mutate(
    ymin = as.numeric(cat) - 0.5,
    ymax = as.numeric(cat) + 0.5
  )
bg_even <- dplyr::filter(bg, row_number() %% 2 == 0)
bg_odd  <- dplyr::filter(bg, row_number() %% 2 == 1)


# Create the single, combined plot
p_final <- ggplot(emm_clean,
       aes(x = solution, y = mean,
           color = job_category,
           shape = job_category,
           group = job_category)) +
# It's important that these rectangles are above the points and
# errors bars, so they'll the the bottom layer on the plot
  geom_rect(data = bg_odd,
            aes(xmin = ymin, xmax = ymax, ymin = -Inf, ymax = Inf),
            inherit.aes = FALSE, fill = "#f8f8f8", color = NA) +
  geom_rect(data = bg_even,
            aes(xmin = ymin, xmax = ymax, ymin = -Inf, ymax = Inf),
            inherit.aes = FALSE, fill = "#e6e6e6", color = NA) +
  geom_hline(yintercept = seq(1, 3, 0.5), color = "gray90") +
  geom_errorbar(aes(ymin = lwr, ymax = upr),
                width = 0.2,
                linewidth = 0.5,
                position = pd) +
  geom_point(size = 5, position = pd) +
```

```
  scale_shape_manual(values = c(16, 17, 15, 18)) +
  scale_x_discrete(limits = interleaved, breaks = soln_levels) +
  ylim(c(1, 3)) +
  labs(
    title = "Estimated Mean Rating by Job Category",
    x = NULL,
    y = "Estimated mean rating"
  ) +
  coord_flip() +
  theme(
    plot.title = element_text(size = 26, hjust = 0, face = "bold"),
    axis.text.x = element_text(size = 20),
    axis.text.y = element_text(size = 20),
    axis.title.x = element_text(size = 24),
    panel.background = element_blank(),
    #panel.grid.major.x = element_line(linetype = "solid", color = "gray90"),
    #panel.grid.major.y = element_line(linetype = "dashed", color = "gray95"),
    plot.margin = unit(c(0.3, 0.3, 0.3, 0.3), "cm"),
    legend.title = element_blank(),
    legend.text=element_text(size=20)
  )

p_final
```

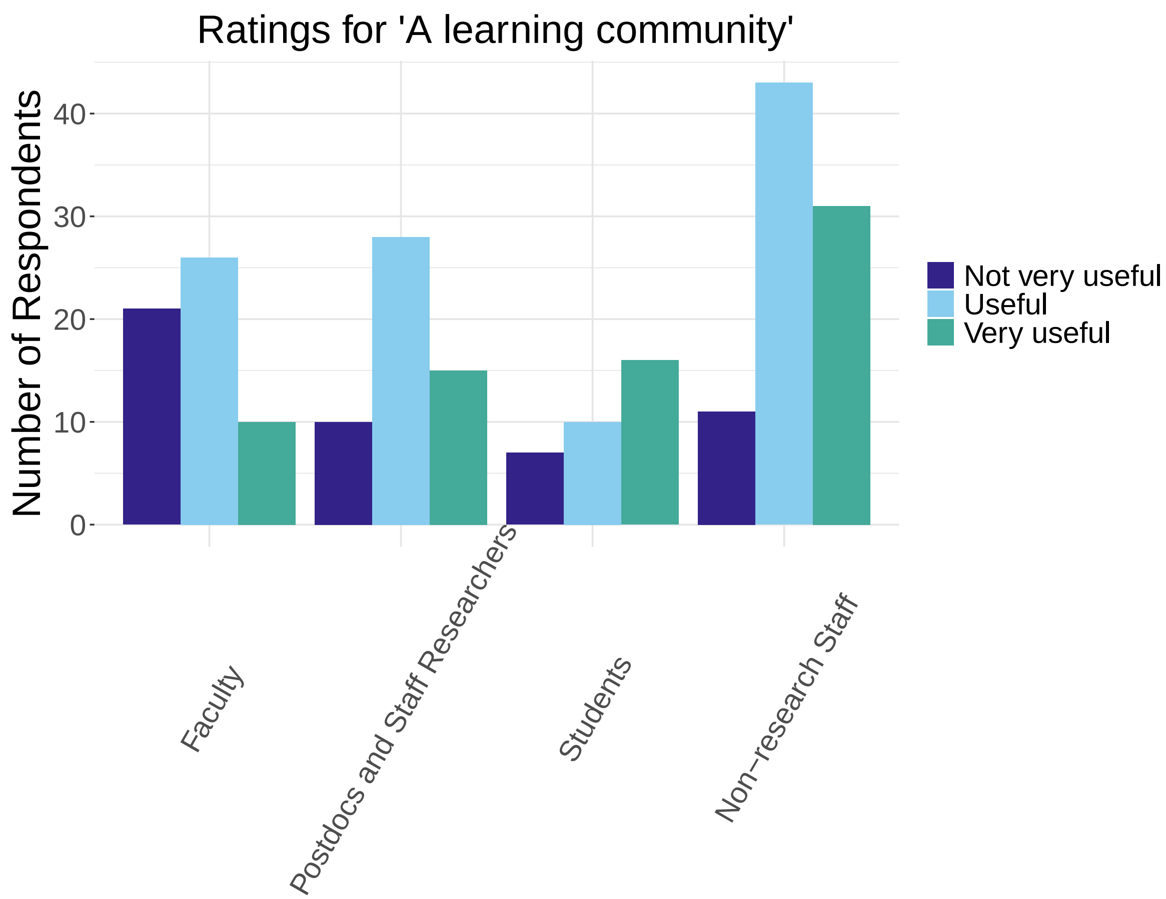



```
save_plot("solns_points2.tiff", 16, 10, p=p_final)
```

Sanity checking: bar plots

I find it very surprising that "a learning community" ranked so low. Let's look at the rating distribution for each job category, for this solution, from the observed sample.

```
learning_ratings <- grouped_bar_chart(
  df = subset(combined, solution=="A learning community"),
  x_var = "job_category",
  fill_var = "utility",
  title = "Ratings for 'A learning community'")
learning_ratings
```




```
save_plot("solns_learning_comm.tiff", 12, 10, p=learning_ratings)
```

Ok, it's a messy plot but whatever. It shows that a lot of non-research staff selected "useful" or "very useful".

Pairwise comparisons and p-values

```
emm_job <- emmeans(fit1b, ~ job_category * solution, mode = "mean.class")
by_job <- summary(
  pairs(emm_job, by = "job_category"),
  infer = TRUE # infer CIs
)
head(by_job)
```

```
 contrast                                      job_category estimate    SE  df
 Computing environments - Publicity            Faculty         0.236 0.143 Inf
 Computing environments - Containerization     Faculty         0.379 0.142 Inf
 Computing environments - Documentation help   Faculty         0.433 0.140 Inf
 Computing environments - A learning community Faculty         0.562 0.135 Inf
 Computing environments - Event planning       Faculty         0.655 0.138 Inf
 Computing environments - Mentoring programs   Faculty         0.689 0.136 Inf
 asymp.LCL asymp.UCL z.ratio p.value
   -0.2299     0.702   1.656  0.8878
   -0.0856     0.844   2.666  0.2433
   -0.0251     0.891   3.089  0.0849
    0.1203     1.003   4.158  0.0019
    0.2029     1.107   4.735  0.0001
    0.2447     1.133   5.068  <.0001

Confidence level used: 0.95
Conf-level adjustment: tukey method for comparing a family of 12 estimates
P value adjustment: tukey method for comparing a family of 12 estimates
```

Here we look at pairwise contrasts by solution. These data are sort-of-kind-of corroborated below with a Kruskal-Wallis test.

```
emm_soln <- emmeans(fit1b, ~ job_category * solution, mode = "mean.class")
by_soln <- summary(
  pairs(emm_soln, by = "solution"),
  infer = TRUE # infer CIs
)

head(by_soln)
```

```
 contrast                                             solution
 Faculty - Postdocs and Staff Researchers             Computing environments
 Faculty - Students                                   Computing environments
 Faculty - (Non-research Staff)                       Computing environments
 Postdocs and Staff Researchers - Students            Computing environments
 Postdocs and Staff Researchers - (Non-research Staff) Computing environments
 Students - (Non-research Staff)                      Computing environments
 estimate    SE  df asymp.LCL asymp.UCL z.ratio p.value
   0.0162 0.169 Inf    -0.419    0.4510   0.096  0.9997
  -0.4426 0.146 Inf    -0.818   -0.0676  -3.032  0.0130
   0.0287 0.151 Inf    -0.360    0.4178   0.189  0.9976
  -0.4588 0.151 Inf    -0.846   -0.0720  -3.047  0.0124
```

```
   0.0125 0.156 Inf    -0.388    0.4131   0.080  0.9998
   0.4713 0.130 Inf     0.136    0.8061   3.616  0.0017

Confidence level used: 0.95
Conf-level adjustment: tukey method for comparing a family of 4 estimates
P value adjustment: tukey method for comparing a family of 4 estimates
```

Let's glance at the significant differences.

```r
# Because there are so many significant comparisons,
# let's be stringent
sig_by_job <- subset(by_job, p.value < 0.0005)
sig_by_job
```

```
                                            contrast
5          Computing environments - Event planning
6      Computing environments - Mentoring programs
20               Publicity - Sustainability grants
29         Containerization - Sustainability grants
30          Containerization - Help finding funding
37       Documentation help - Sustainability grants
38        Documentation help - Help finding funding
44     A learning community - Sustainability grants
45      A learning community - Help finding funding
50           Event planning - Sustainability grants
51            Event planning - Help finding funding
55       Mentoring programs - Sustainability grants
56        Mentoring programs - Help finding funding
59                Education - Sustainability grants
60                 Education - Help finding funding
62            Legal support - Sustainability grants
63             Legal support - Help finding funding
64  Industry partnerships - Sustainability grants
95         Containerization - Sustainability grants
96          Containerization - Help finding funding
110    A learning community - Sustainability grants
116          Event planning - Sustainability grants
117           Event planning - Help finding funding
121      Mentoring programs - Sustainability grants
122       Mentoring programs - Help finding funding
125               Education - Sustainability grants
128           Legal support - Sustainability grants
```

```
130 Industry partnerships - Sustainability grants
182        Event planning - Sustainability grants
200     Computing environments - Containerization
203       Computing environments - Event planning
218             Publicity - Sustainability grants
227      Containerization - Sustainability grants
228       Containerization - Help finding funding
248        Event planning - Sustainability grants
249         Event planning - Help finding funding
262 Industry partnerships - Sustainability grants
                      job_category   estimate        SE  df  asymp.LCL
5                          Faculty  0.6549404 0.1383278 Inf  0.2028847
6                          Faculty  0.6888481 0.1359182 Inf  0.2446670
20                         Faculty -0.6897175 0.1246919 Inf -1.0972112
29                         Faculty -0.8326377 0.1244847 Inf -1.2394542
30                         Faculty -0.7008841 0.1312171 Inf -1.1297021
37                         Faculty -0.8863025 0.1220734 Inf -1.2852387
38                         Faculty -0.7545489 0.1289687 Inf -1.1760190
44                         Faculty -1.0153942 0.1162150 Inf -1.3951852
45                         Faculty -0.8836407 0.1237773 Inf -1.2881454
50                         Faculty -1.1084421 0.1195653 Inf -1.4991821
51                         Faculty -0.9766886 0.1267895 Inf -1.3910373
55                         Faculty -1.1423498 0.1164505 Inf -1.5229105
56                         Faculty -1.0105963 0.1240400 Inf -1.4159595
59                         Faculty -1.0569497 0.1189241 Inf -1.4455941
60                         Faculty -0.9251962 0.1264674 Inf -1.3384922
62                         Faculty -0.8759528 0.1191070 Inf -1.2651948
63                         Faculty -0.7441992 0.1260213 Inf -1.1560374
64                         Faculty -0.6954577 0.1233462 Inf -1.0985536
95  Postdocs and Staff Researchers -0.8453910 0.1222004 Inf -1.2447424
96  Postdocs and Staff Researchers -0.7123321 0.1286174 Inf -1.1326544
110 Postdocs and Staff Researchers -0.6462254 0.1204395 Inf -1.0398220
116 Postdocs and Staff Researchers -0.7792098 0.1252404 Inf -1.1884959
117 Postdocs and Staff Researchers -0.6461509 0.1314915 Inf -1.0758656
121 Postdocs and Staff Researchers -0.7701944 0.1265056 Inf -1.1836151
122 Postdocs and Staff Researchers -0.6371354 0.1325076 Inf -1.0701709
125 Postdocs and Staff Researchers -0.6385922 0.1243440 Inf -1.0449489
128 Postdocs and Staff Researchers -0.6798460 0.1260943 Inf -1.0919226
130 Postdocs and Staff Researchers -0.5735799 0.1276135 Inf -0.9906213
182                       Students -0.7519809 0.1616113 Inf -1.2801274
200              Non-research Staff  0.5332636 0.1160426 Inf  0.1540358
203              Non-research Staff  0.5309440 0.1161705 Inf  0.1512982
218              Non-research Staff -0.6799672 0.1147994 Inf -1.0551319
```

```
227            Non-research Staff -0.7053335 0.1153851 Inf -1.0824125
228            Non-research Staff -0.5485665 0.1184652 Inf -0.9357112
248            Non-research Staff -0.7030140 0.1149538 Inf -1.0786835
249            Non-research Staff -0.5462469 0.1181389 Inf -0.9323254
262            Non-research Staff -0.5829469 0.1177207 Inf -0.9676586
     asymp.UCL   z.ratio      p.value
5    1.1069962  4.734699 1.388367e-04
6    1.1330292  5.068109 2.590106e-05
20  -0.2822239 -5.531373 2.076395e-06
29  -0.4258211 -6.688674 1.485164e-09
30  -0.2720661 -5.341408 6.000927e-06
37  -0.4873662 -7.260408 2.554301e-11
38  -0.3330789 -5.850638 3.215137e-07
44  -0.6356033 -8.737207 9.514611e-14
45  -0.4791360 -7.138956 6.212852e-11
50  -0.7177022 -9.270598 1.266764e-13
51  -0.5623399 -7.703227 9.453549e-13
55  -0.7617891 -9.809746 1.245670e-13
56  -0.6052330 -8.147340 1.683098e-13
59  -0.6683053 -8.887600 8.648637e-14
60  -0.5119001 -7.315688 1.697409e-11
62  -0.4867107 -7.354337 1.273992e-11
63  -0.3323611 -5.905344 2.311823e-07
64  -0.2923618 -5.638257 1.124693e-06
95  -0.4460396 -6.918070 3.021001e-10
96  -0.2920098 -5.538379 1.995309e-06
110 -0.2526288 -5.365562 5.254050e-06
116 -0.3699237 -6.221714 3.237886e-08
117 -0.2164362 -4.914013 5.710042e-05
121 -0.3567737 -6.088225 7.511344e-08
122 -0.2041000 -4.808292 9.681196e-05
125 -0.2322355 -5.135689 1.817081e-05
128 -0.2677695 -5.391570 4.550438e-06
130 -0.1565385 -4.494665 4.321159e-04
182 -0.2238344 -4.653021 2.057479e-04
200  0.9124913  4.595411 2.703628e-04
203  0.9105898  4.570384 3.040797e-04
218 -0.3048024 -5.923092 2.075892e-07
227 -0.3282546 -6.112865 6.439254e-08
228 -0.1614218 -4.630613 2.289050e-04
248 -0.3273444 -6.115621 6.329075e-08
249 -0.1601684 -4.623767 2.364616e-04
262 -0.1982351 -4.951948 4.711072e-05
```

```r
sig_by_soln <- subset(by_soln, p.value < 0.05)
sig_by_soln
```

```
                                          contrast                solution
2                               Faculty - Students Computing environments
4              Postdocs and Staff Researchers - Students Computing environments
6                     Students - (Non-research Staff) Computing environments
11 Postdocs and Staff Researchers - (Non-research Staff)               Publicity
12                    Students - (Non-research Staff)               Publicity
18                    Students - (Non-research Staff)        Containerization
26                              Faculty - Students   A learning community
27                     Faculty - (Non-research Staff)   A learning community
38                              Faculty - Students     Mentoring programs
39                     Faculty - (Non-research Staff)     Mentoring programs
44                              Faculty - Students               Education
56                              Faculty - Students  Industry partnerships
60                    Students - (Non-research Staff)  Industry partnerships
63                     Faculty - (Non-research Staff)  Sustainability grants
66                    Students - (Non-research Staff)  Sustainability grants
72                    Students - (Non-research Staff)   Help finding funding
     estimate        SE  df   asymp.LCL   asymp.UCL   z.ratio      p.value
2  -0.4425989 0.1459660 Inf -0.81759018 -0.06760760 -3.032205 1.297267e-02
4  -0.4588231 0.1505653 Inf -0.84563008 -0.07201615 -3.047337 1.236403e-02
6   0.4712768 0.1303217 Inf  0.13647631  0.80607726  3.616258 1.700432e-03
11  0.5592622 0.1551499 Inf  0.16067722  0.95784717  3.604658 1.776658e-03
12  0.6995423 0.1761166 Inf  0.24709308  1.15199158  3.972040 4.142473e-04
18  0.5475497 0.1905313 Inf  0.05806868  1.03703070  2.873804 2.113462e-02
26 -0.5588410 0.1918513 Inf -1.05171317 -0.06596874 -2.912885 1.878472e-02
27 -0.4884963 0.1487836 Inf -0.87072610 -0.10626660 -3.283268 5.661492e-03
38 -0.7545348 0.1853409 Inf -1.23068154 -0.27838811 -4.071064 2.732612e-04
39 -0.4148289 0.1521663 Inf -0.80574893 -0.02390883 -2.726155 3.249277e-02
44 -0.6623738 0.1901221 Inf -1.15080345 -0.17394418 -3.483940 2.780828e-03
56 -0.4780319 0.1798083 Inf -0.93996518 -0.01609858 -2.658564 3.924501e-02
60  0.6756306 0.1673319 Inf  0.24574967  1.10551161  4.037668 3.147720e-04
63  0.3101096 0.1126268 Inf  0.02076775  0.59945146  2.753426 3.006602e-02
66  0.4717569 0.0955300 Inf  0.22633727  0.71717647  4.938311 4.698058e-06
72  0.5040808 0.1238873 Inf  0.18581024  0.82235131  4.068864 2.758288e-04
```

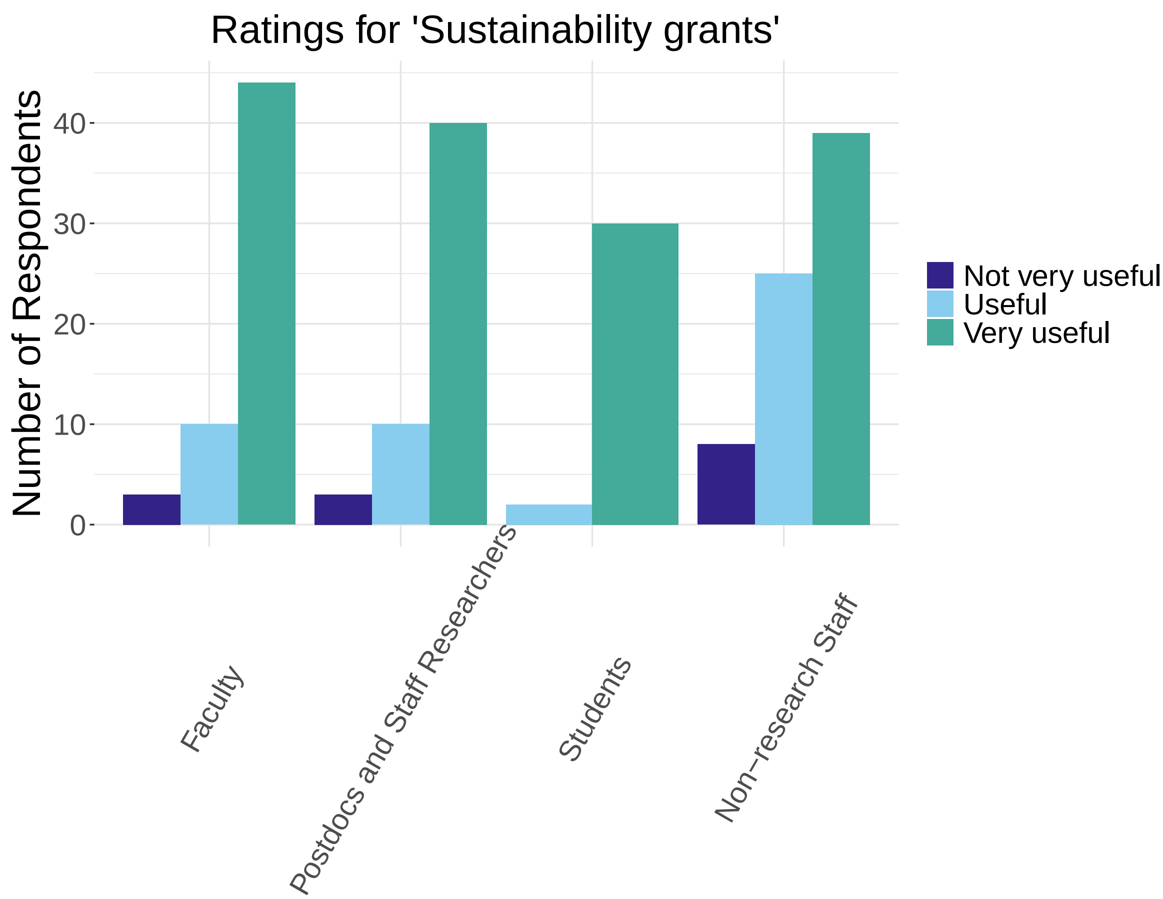
Okay, so here, the "estimate" column shows the difference in estimated marginal means for the two levels of interest, holding the other factor level constant (of my two factors, job and solution). So when the contrast is Computing environments vs. A learning community, the job_category is Faculty, and the estimate is 0.57, this indicates that the difference between the estimates of faculty's average rating of computer environments and their average rating of a learning community is 0.56, on a three-point scale.

```
subset(
  summary(emm),
  job_category == "Faculty" & solution == "Computing environments"
)$mean.class -
  subset(
    summary(emm),
    job_category == "Faculty" & solution == "A learning community"
  )$mean.class
```

```
[1] 0.5618926
```

Sustainability grants and Finding funding show up frequently as being significantly higher than some of the other solutions. Let's plot the distributions of responses for sustainability grants, as a sanity check.

```
grant_ratings <- grouped_bar_chart(
  df = subset(combined, solution=="Sustainability grants"),
  x_var = "job_category",
  fill_var = "utility",
  title = "Ratings for 'Sustainability grants'")
grant_ratings
```




```
save_plot("solns_grants.tiff", 12, 10, p=grant_ratings)
```

Kruskal-Wallis test for ranking differences between groups

Non-parametric corroboration of the extent of disagreement between groups. Whereas above, we tested for differences in mean ratings, here we are testing for differences in the distributions of ratings for each solution.

```
combined2 <- combined %>%
  mutate(
    utility_score = recode(
      utility,
      "Non-applicable" = 0L,
```

```
      "Not very useful" = 0L,
      "Useful" = 1L,
      "Very useful" = 2L
    )
  )

kw_results <- sapply(split(combined2, combined2$solution), function(df) {
  kruskal.test(utility_score ~ job_category, data = df)$p.value
})

p_adj_kw <- p.adjust(kw_results, "holm")

p_adj_kw < 0.05
```

```
Computing environments              Publicity        Containerization
                 FALSE                   TRUE                   FALSE
    Documentation help   A learning community          Event planning
                 FALSE                   TRUE                   FALSE
     Mentoring programs              Education           Legal support
                  TRUE                  FALSE                   FALSE
  Industry partnerships  Sustainability grants    Help finding funding
                 FALSE                   TRUE                    TRUE
```

```
sum(p_adj_kw < 0.05)
```

```
[1] 5
```

Hm. The results are a little surprising. Only five solutions are "divisive" according to this test. But maybe it's not surprising, if the whole point of doing ordinal regression is that it's more sensitive than a non-parametric test.

Session Info

```
sessionInfo()
```

```
R version 4.4.2 (2024-10-31)
Platform: aarch64-apple-darwin20
```

```
Running under: macOS Sequoia 15.6.1

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib;  

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

time zone: America/Los_Angeles
tzcode source: internal

attached base packages:
[1] tools     grid      stats     graphics  grDevices datasets  utils
[8] methods   base

other attached packages:
 [1] treemapify_2.5.6      tidyr_1.3.1           svglite_2.2.1
 [4] stringr_1.5.1         scales_1.4.0          readr_2.1.5
 [7] pwr_1.3-0             patchwork_1.3.2       ordinal_2023.12-4.1
[10] lme4_1.1-37           Matrix_1.7-1          languageserver_0.3.16
[13] here_1.0.1            gtools_3.9.5          ggforce_0.5.0
[16] FSA_0.10.0            fpc_2.2-13            forcats_1.0.0
[19] factoextra_1.0.7      ggplot2_3.5.2         emmeans_1.11.2
[22] dplyr_1.1.4           corrplot_0.95         ComplexHeatmap_2.22.0
[25] cluster_2.1.8.1       BiocManager_1.30.26

loaded via a namespace (and not attached):
 [1] Rdpack_2.6.4        rlang_1.1.6         magrittr_2.0.3
 [4] clue_0.3-66         GetoptLong_1.0.5    matrixStats_1.5.0
 [7] compiler_4.4.2      flexmix_2.3-20      systemfonts_1.2.3
[10] png_0.1-8           callr_3.7.6         vctrs_0.6.5
[13] pkgconfig_2.0.3     shape_1.4.6.1       crayon_1.5.3
[16] fastmap_1.2.0       labeling_0.4.3      utf8_1.2.6
[19] rmarkdown_2.29      ggfittext_0.10.2    tzdb_0.5.0
[22] ps_1.9.1            nloptr_2.2.1        purrr_1.1.0
[25] xfun_0.53           modeltools_0.2-24   jsonlite_2.0.0
[28] tweenr_2.0.3        parallel_4.4.2      prabclus_2.3-4
[31] R6_2.6.1            stringi_1.8.7       RColorBrewer_1.1-3
[34] boot_1.3-31         diptest_0.77-2      numDeriv_2016.8-1.1
[37] estimability_1.5.1  Rcpp_1.1.0          iterators_1.0.14
[40] knitr_1.50          IRanges_2.40.1      splines_4.4.2
[43] nnet_7.3-19         tidyselect_1.2.1    yaml_2.3.10
```

```
[46] doParallel_1.0.17   codetools_0.2-20    processx_3.8.6
[49] lattice_0.22-6      tibble_3.3.0        withr_3.0.2
[52] evaluate_1.0.4      polyclip_1.10-7     xml2_1.4.0
[55] circlize_0.4.16     mclust_6.1.1        kernlab_0.9-33
[58] pillar_1.11.0       renv_1.1.5          foreach_1.5.2
[61] stats4_4.4.2        reformulas_0.4.1    generics_0.1.4
[64] rprojroot_2.1.1     S4Vectors_0.44.0    hms_1.1.3
[67] minqa_1.2.8         xtable_1.8-4        class_7.3-22
[70] glue_1.8.0          robustbase_0.99-4-1 mvtnorm_1.3-3
[73] rbibutils_2.3       colorspace_2.1-1    nlme_3.1-166
[76] cli_3.6.5           textshaping_1.0.1   gtable_0.3.6
[79] DEoptimR_1.1-4      digest_0.6.37       BiocGenerics_0.52.0
[82] ucminf_1.2.2        ggrepel_0.9.6       rjson_0.2.23
[85] farver_2.1.2        htmltools_0.5.8.1   lifecycle_1.0.4
[88] GlobalOptions_0.1.2 MASS_7.3-61
```