```r
# Installing libraries
library(TAM) # for Rasch modeling
library(WrightMap) # to build item-person (Wright) maps
library(lsr)  # to calculate Cohen's d
library(DescTools)  # to calculate eta-squared
library(Rmisc) # to calculate summary statistics
library(tidyverse) # for plotting
library(cowplot) # for image building
library(magick) # for image processing
library(knitr) # for knitting



# Create logo banner for plots
earthlab_orig <- image_read(path = "earth-lab-logo-white.png") %>%
    image_scale("x80")

twitter_orig <- image_read(path = "plot-footer-twitter.png") %>%
    image_scale("x70")

black_banner <- image_read(path = "black-banner.png")

earthlab_logo <- image_composite(image_scale(black_banner, "1000x100"), earthlab_orig, o
ffset = "+30+10")
twitter_logo <- image_composite(image_scale(black_banner, "1000x100"), twitter_orig, off
set = "+540+15")

logo <- image_append(image_scale(c(earthlab_logo, twitter_logo)), stack = FALSE)

logo
```
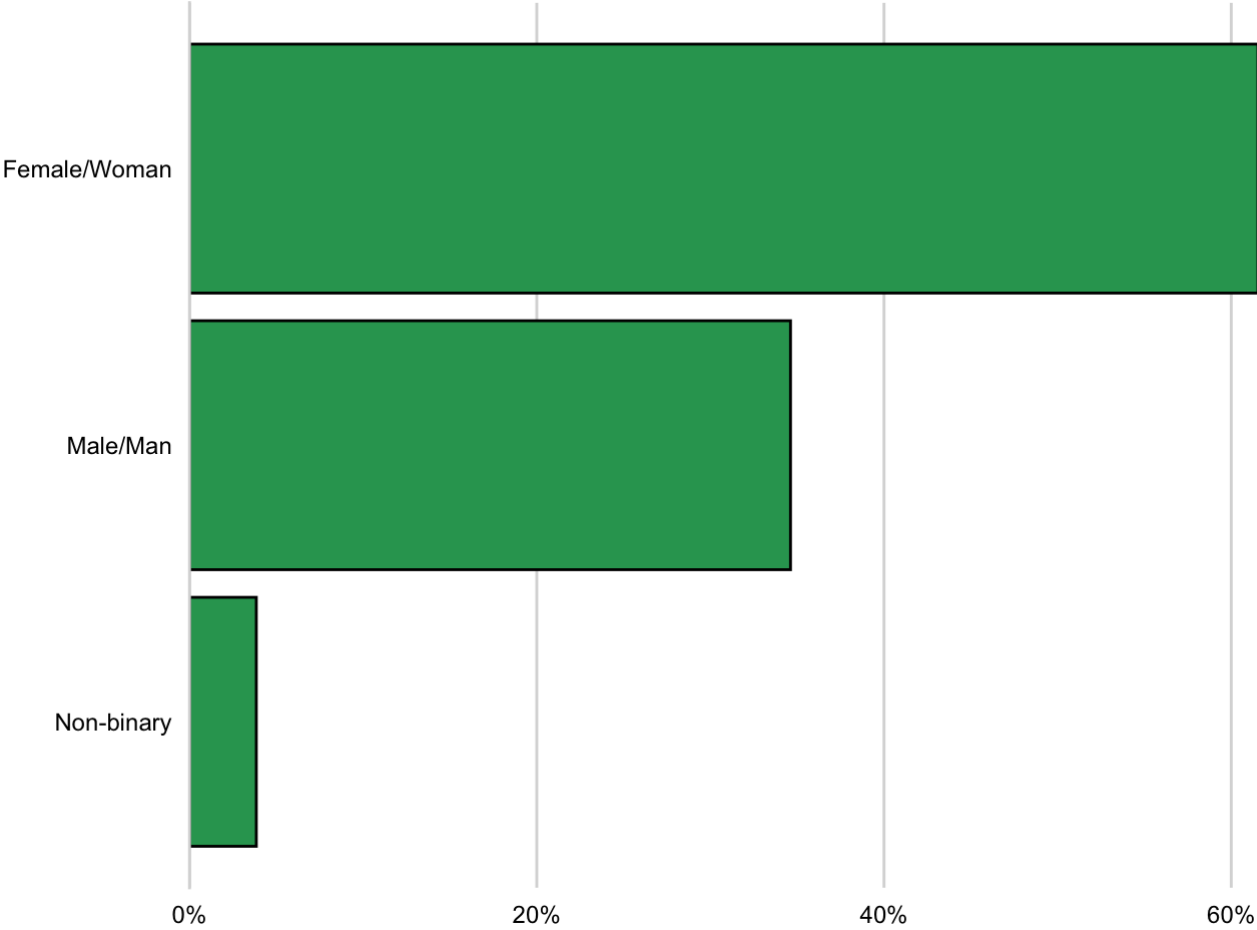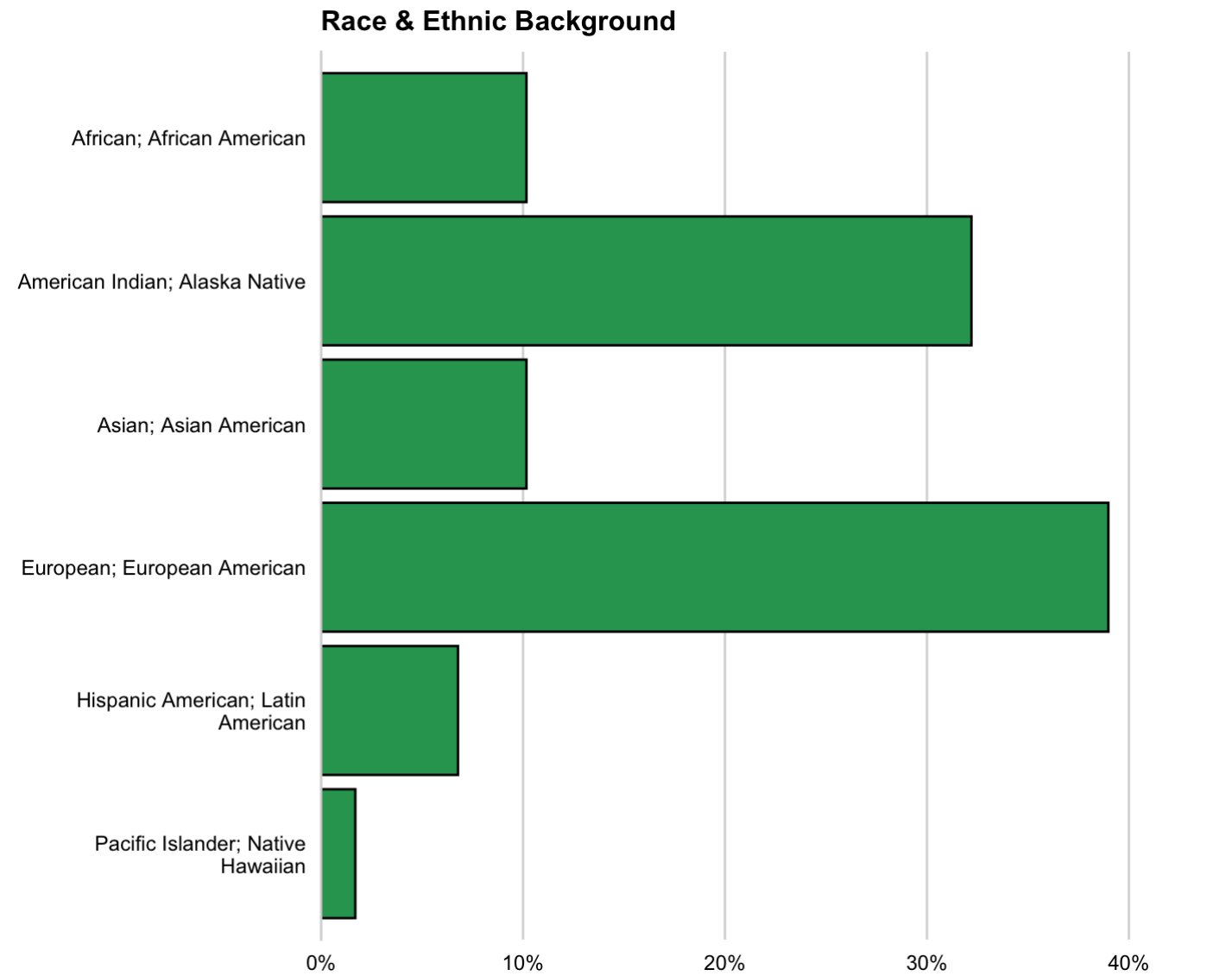

Participant Demographics

A total of 53 consenting participants provided demographic information related to gender, race & ethnicity through a series of items included on the pre-program survey instrument, administered prior to the start of the technical workshops.
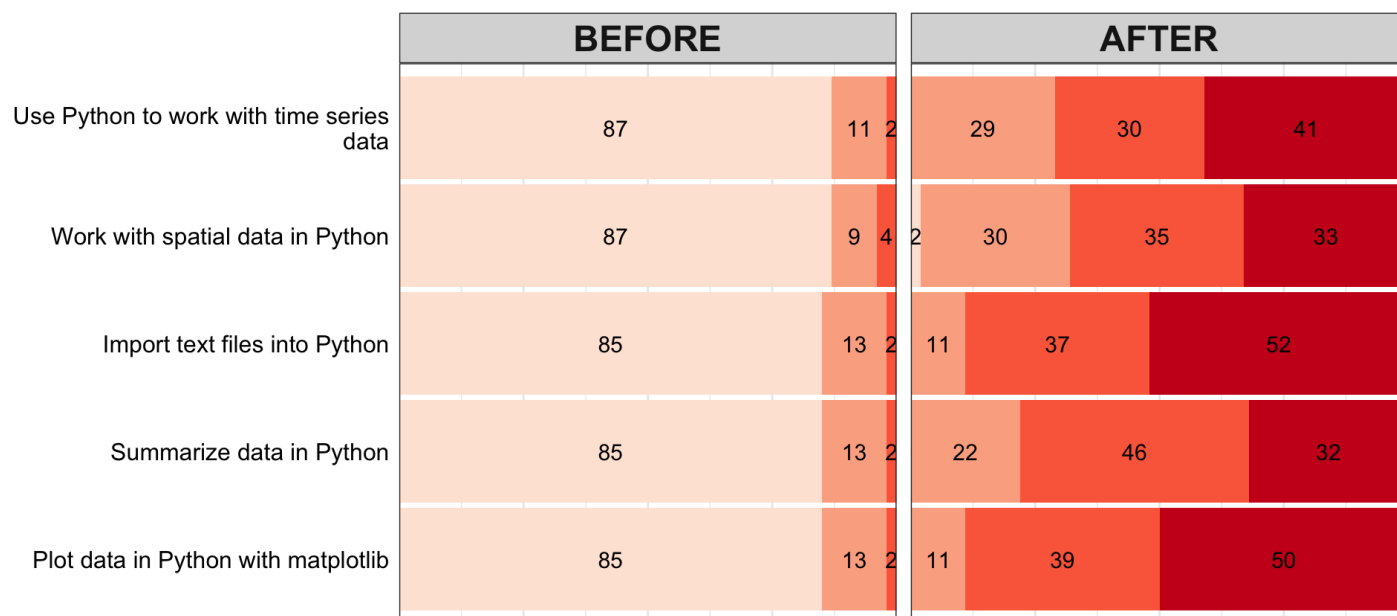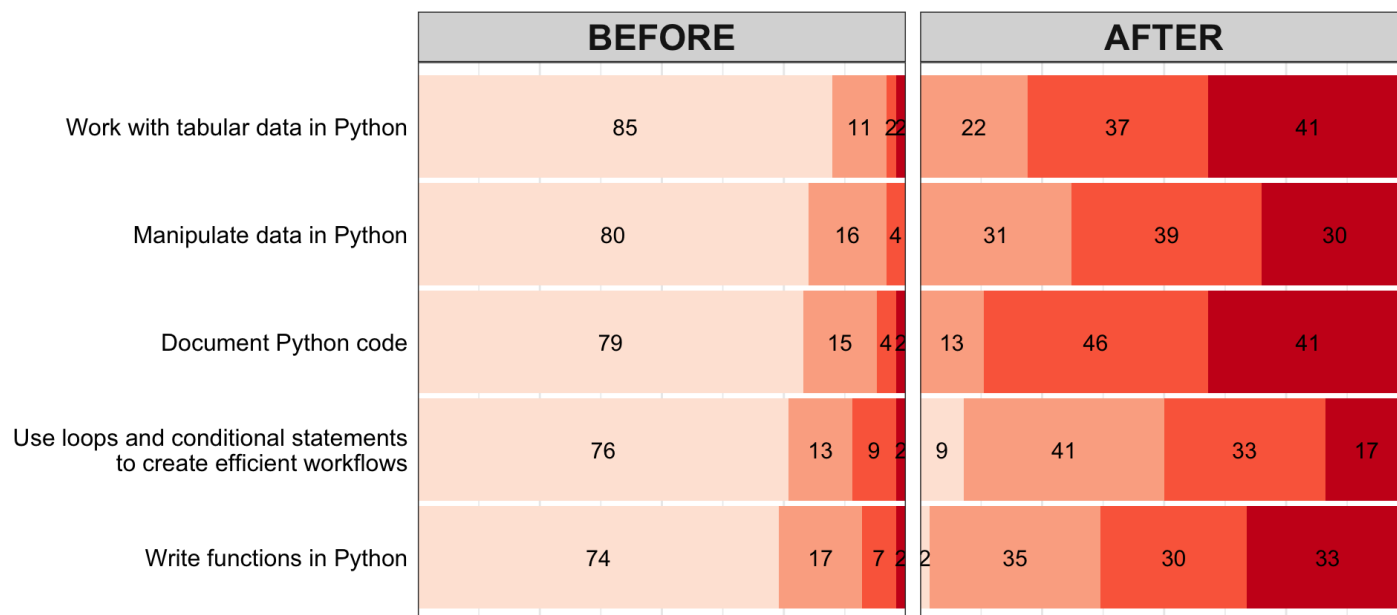
Race & Ethnic Background

Raw Responses (%)

Python Skills

Not Comfortable   **Slightly Comfortable**   **Moderately Comfortable**   **Very Comfortable**

	BEFORE	AFTER
Use Python to work with time series data	87 · 11 · 2	29 · 30 · 41
Work with spatial data in Python	87 · 9 · 4	2 · 30 · 35 · 33
Import text files into Python	85 · 13 · 2	11 · 37 · 52
Summarize data in Python	85 · 13 · 2	22 · 46 · 32
Plot data in Python with matplotlib	85 · 13 · 2	11 · 39 · 50

Not Comfortable   **Slightly Comfortable**   **Moderately Comfortable**   **Very Comfortable**

	BEFORE	AFTER
Work with tabular data in Python	85 · 11 · 2 2	22 · 37 · 41
Manipulate data in Python	80 · 16 · 4	31 · 39 · 30
Document Python code	79 · 15 · 4 2	13 · 46 · 41
Use loops and conditional statements to create efficient workflows	76 · 13 · 9 · 2	9 · 41 · 33 · 17
Write functions in Python	74 · 17 · 7 2	2 · 35 · 30 · 33

Data Science Skills

Legend: Not Confident | Slightly Confident | Moderately Confident | Very Confident

	BEFORE				AFTER			
I can use Jupyter Notebooks	78	13	7	2	2	20	30	48
I can publish my code	76	13	7	4	7	30	35	28
I can write efficient and modular code	67	24	7	2	2	24	57	17
I can document code so that others can easily understand and use it	61	20	17	2	4	7	57	32
I can use scientific programming to work with data	57	30	11	2		13	57	30

Legend: Not Confident | Slightly Confident | Moderately Confident | Very Confident

	BEFORE				AFTER			
I can use remote sensing data	55	24	17	4	9	28	43	20
I can create reproducible workflows	44	37	15	4	2	20	50	28
I can find and access data	28	33	32	7		9	30	61
I can draw analytical conclusions from data	24	37	30	9		9	50	41
I can visualize scientific data	24	37	32	7		15	37	48

Data Science Communication

Data Science Practices

Legend: Not Confident | Slightly Confident | Moderately Confident | Very Confident

BEFORE / AFTER

	BEFORE	AFTER
Use technical science skills (such as tools, instruments, and techniques)	25, 44, 27, 4	9, 58, 33
Figure out how the data that you've downloaded for a project can be used to address your science question	25, 36, 33, 6	2, 8, 46, 44
Find the data that you need to address a science question	25, 33, 29, 13	13, 43, 44
Communicate project results in a blog format	19, 42, 33, 6	4, 17, 46, 33
Describe the results of your data processing as it relates to your science question	18, 42, 27, 13	8, 46, 46

	BEFORE	AFTER
Use scientific language and terminology	15, 31, 41, 13	6, 63, 31
Relate results and explanations to the work of others	10, 42, 33, 15	12, 44, 44
Communicate project results in an oral presentation	9, 35, 31, 25	4, 48, 48
Use scientific literature and/or reports to guide research	8, 31, 40, 21	4, 50, 46
Generate a research question	8, 23, 52, 17	6, 48, 46

Science Identity

Legend: Strongly Disgree | Disgree | Neither Agree nor Disagree | Agree | Strongly Agree

BEFORE / AFTER

Statement	Before (SD, D, N, A, SA)	After (SD, D, N, A, SA)
I have come to think of myself as a 'scientist'	11, 17, 30, 22, 20	2, 9, 28, 24, 37
Being a scientist is an important reflection of who I am	11, 11, 32, 26, 20	7, 26, 26, 41
In general, being a scientist is an important part of my self-image	9, 9, 28, 30, 24	2, 4, 20, 33, 41
I have a strong sense of belonging to the community of scientists	7, 22, 38, 24, 9	9, 35, 32, 24
I feel like I belong in the field of science	7, 17, 30, 24, 22	7, 15, 30, 48
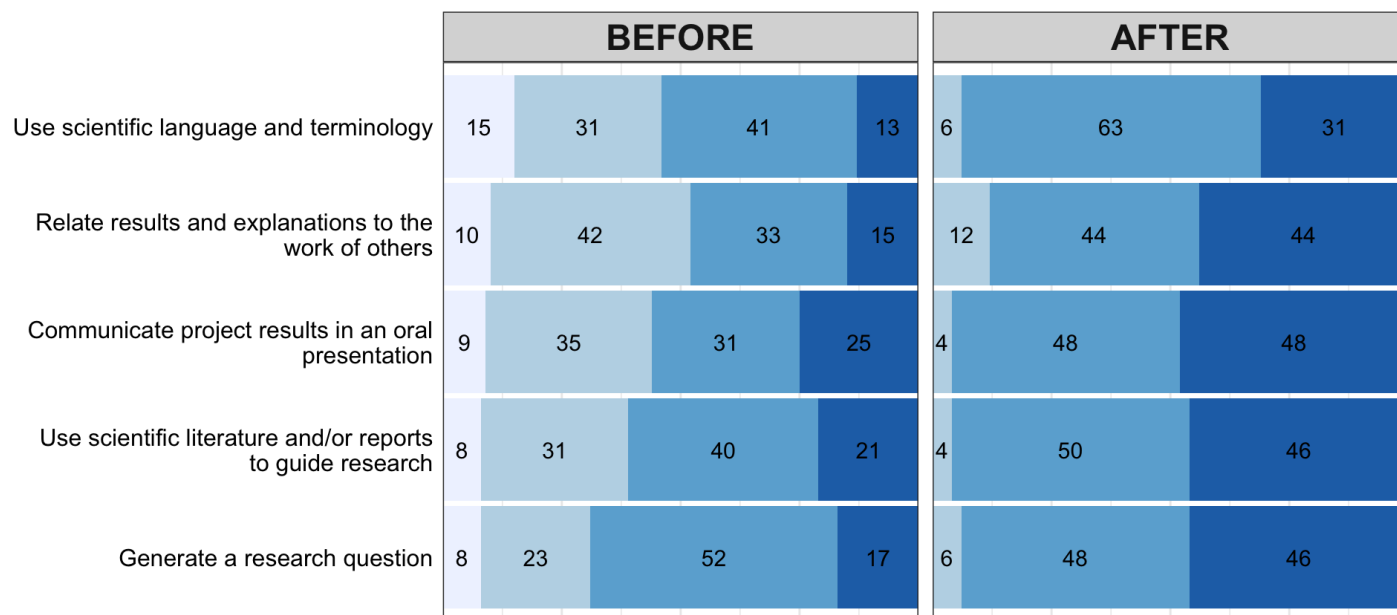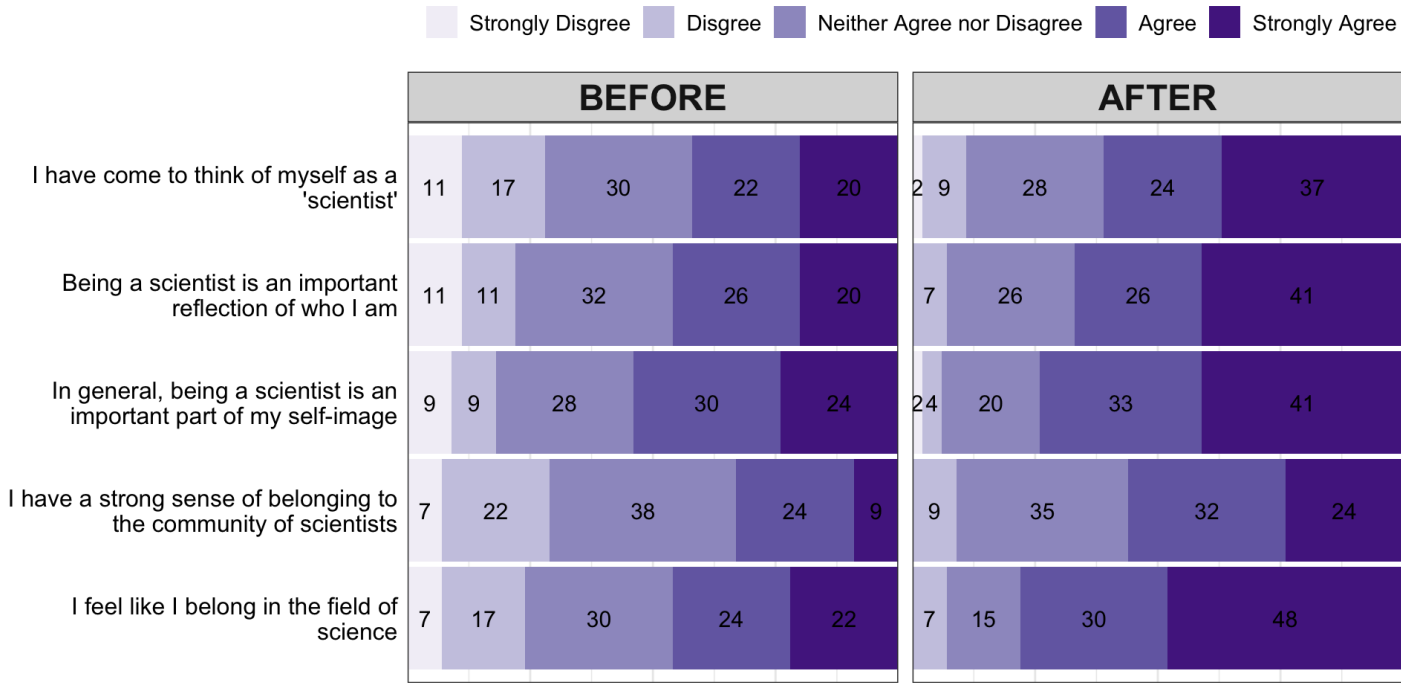
Rasch Modeling (Wright Maps)

Python Skills

Python Skills (BEFORE)

Person Distribution

```
08.4
01.4 | 02.4 | 06.4 | 10.4
09.4
03.4
07.4
05.4 | 08.3
01.3 | 02.3 | 04.4 | 06.3 | 10.3
09.3
03.3
07.3
04.3 | 05.3
08.2
01.2 | 02.2 | 06.2 | 10.2
09.2
03.2
07.2

04.2 | 05.2
```

Logits

Item Distriubtion

Python Skills (AFTER)

Person Distribution

```
08.4
01.4 | 02.4 | 06.4 | 10.4
09.4
03.4
07.4
05.4 | 08.3
01.3 | 02.3 | 04.4 | 06.3 | 10.3
09.3
03.3
07.3
04.3 | 05.3
08.2
01.2 | 02.2 | 06.2 | 10.2
09.2
03.2
07.2
05.2
04.2
```

Logits

Item Distriubtion
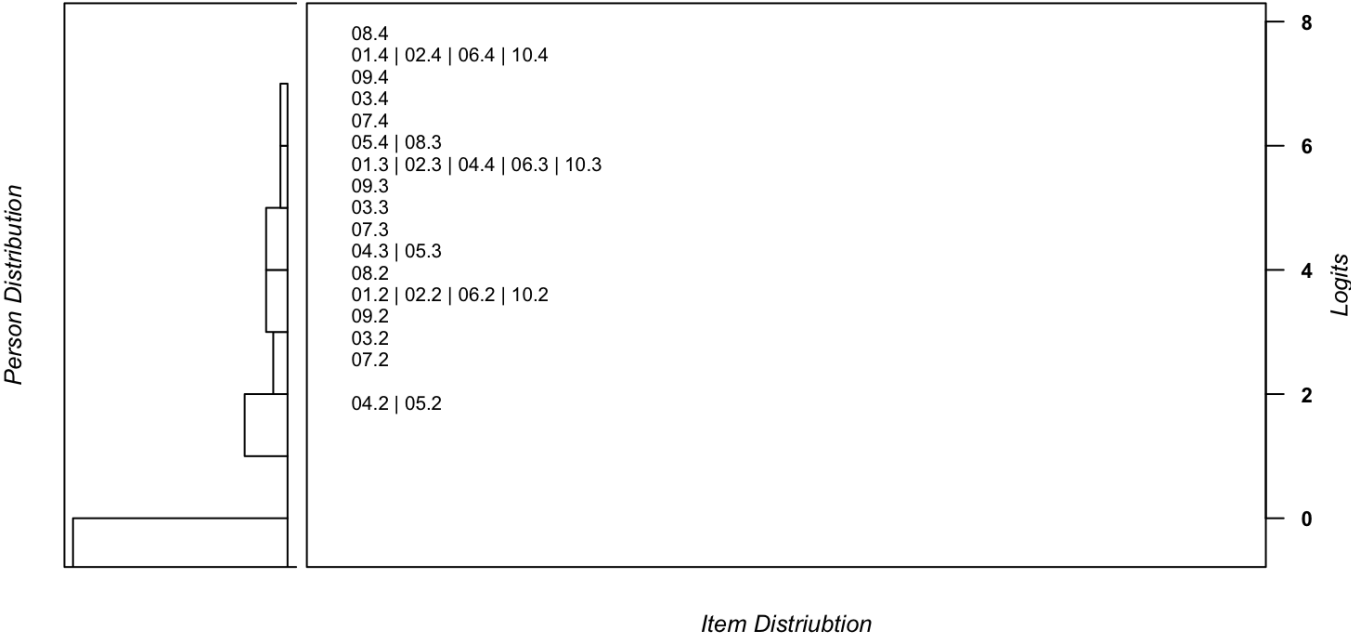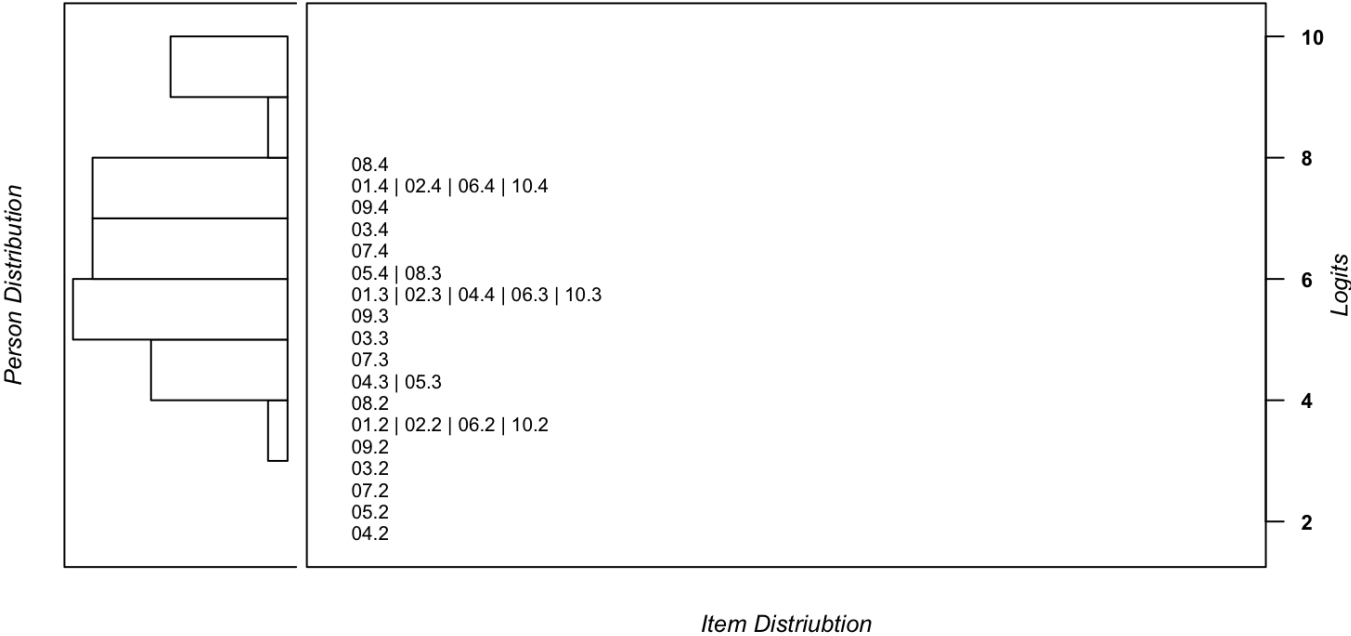
Data Science Skills

Data Science Skills (BEFORE)

Item Distriubtion
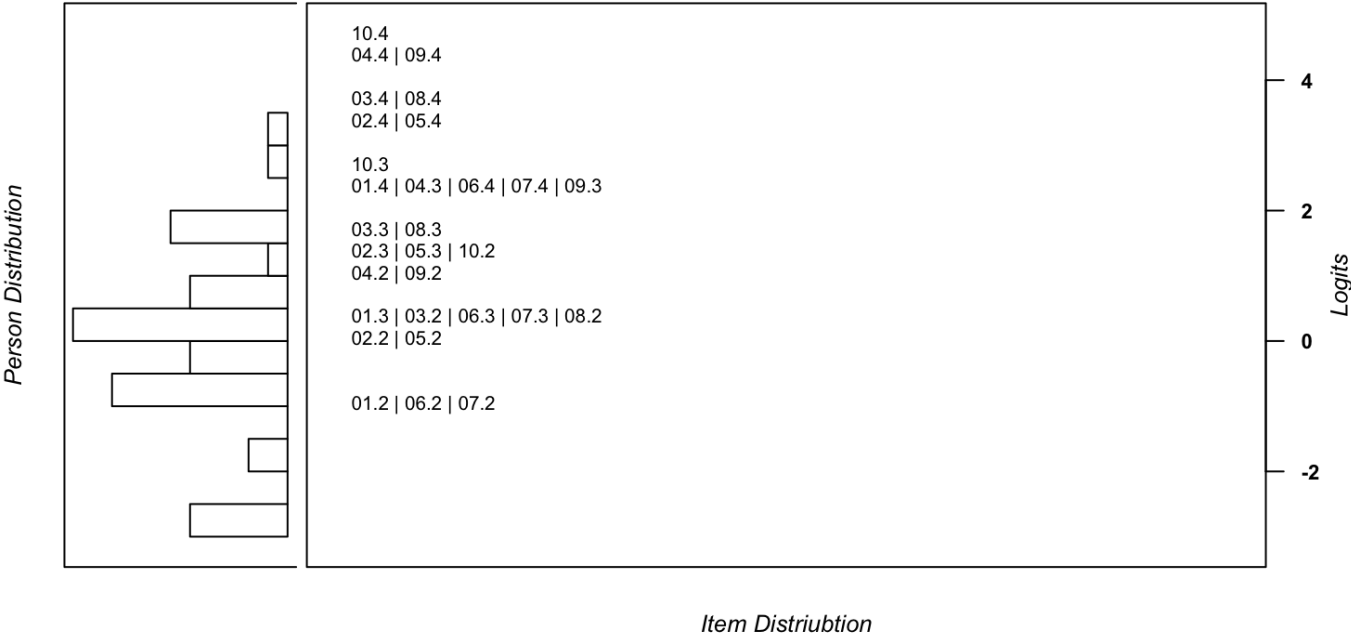
Data Science Skills (AFTER)

Item Distriubtion
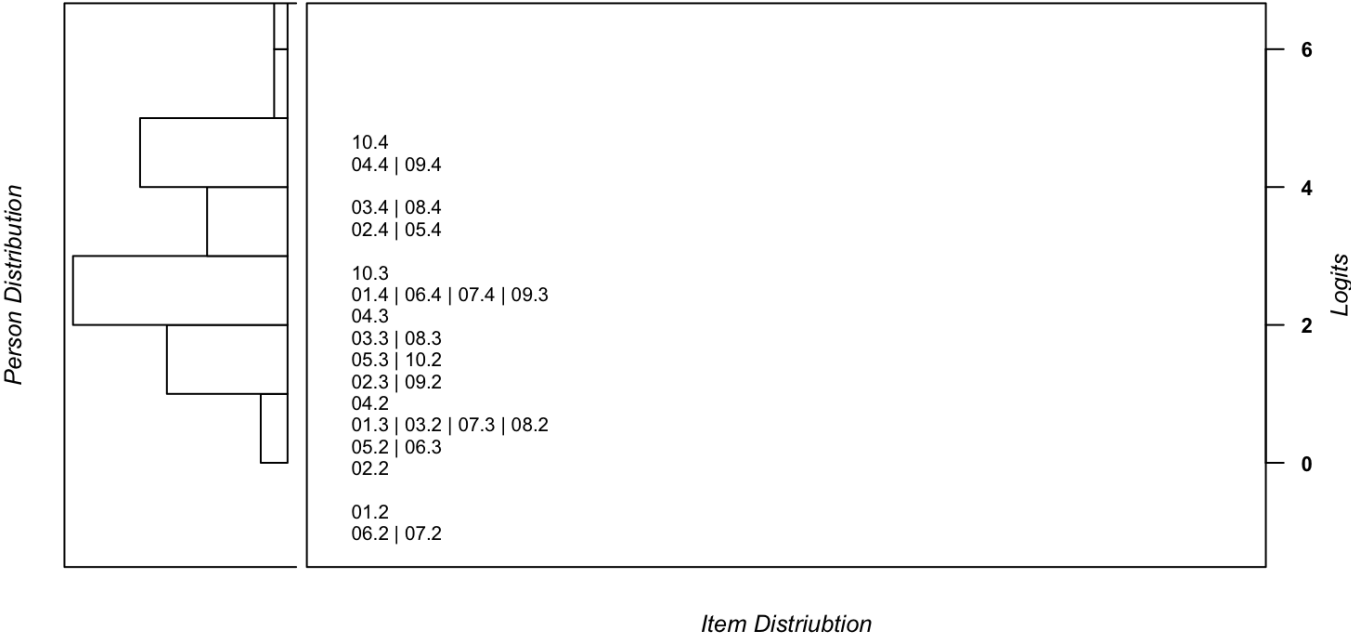
Data Science Communication

Data Science Communication (BEFORE)

Item Distriubtion
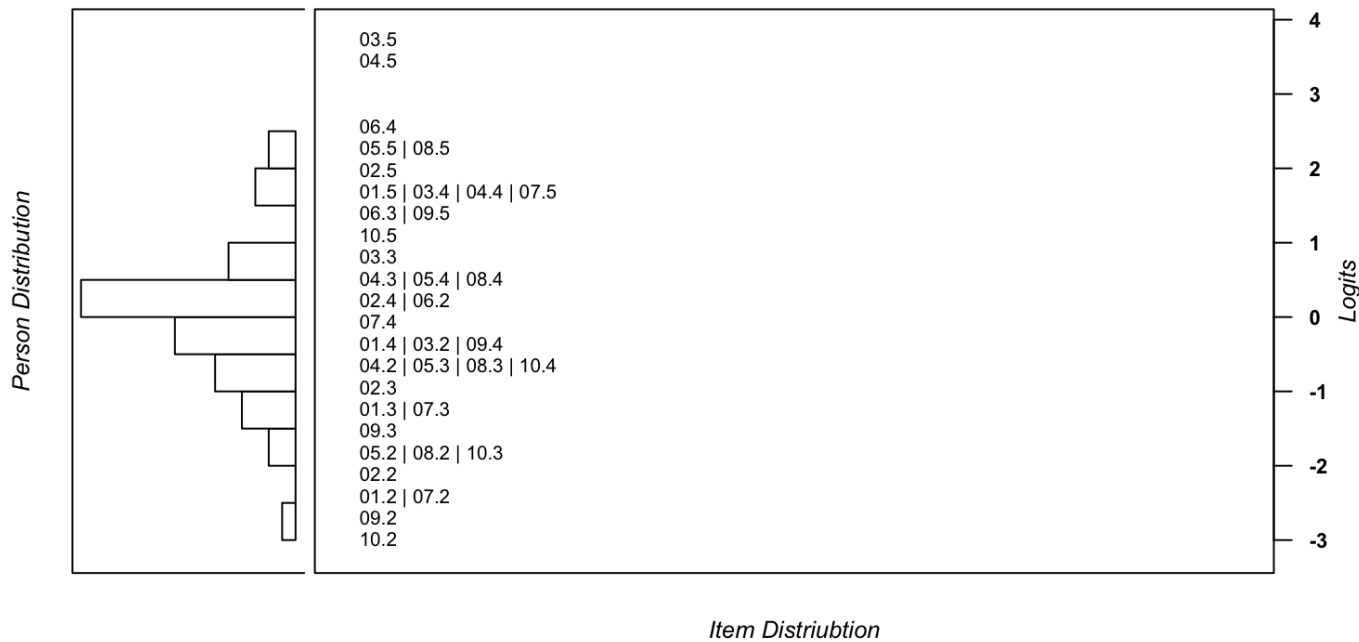
Data Science Communication (AFTER)

Item Distriubtion
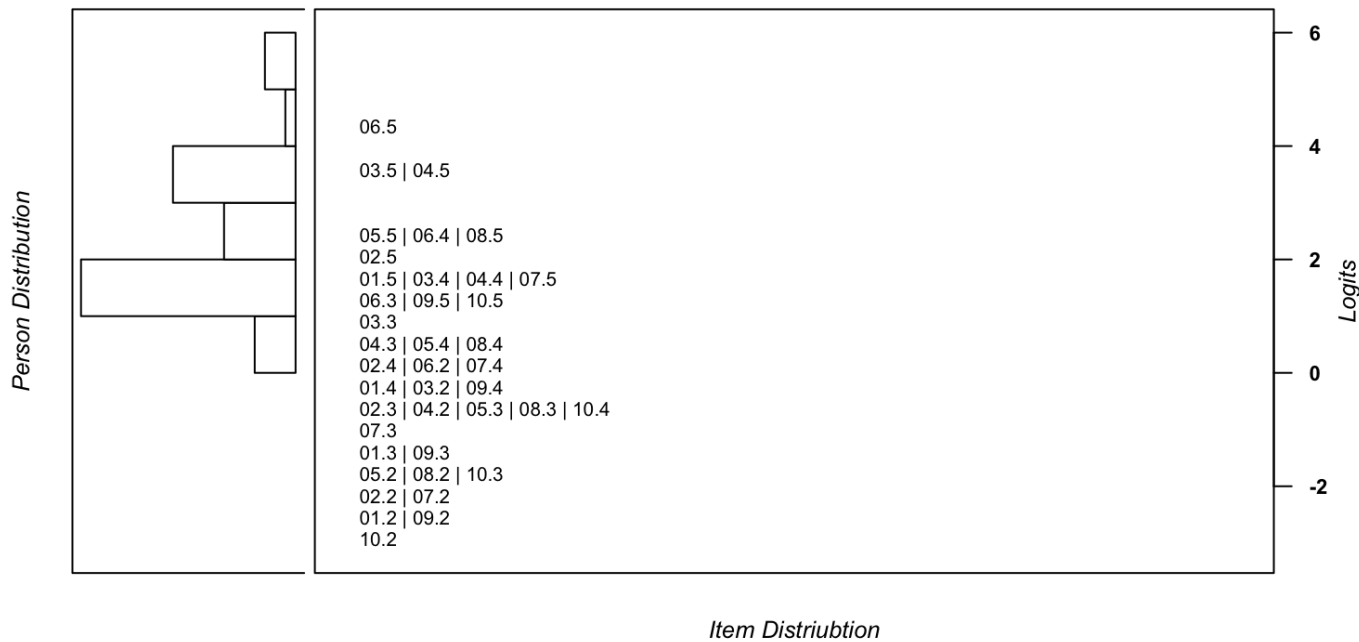
Data Science Practices

Data Science Practices (BEFORE)

Person Distribution

01.4	
05.4	
04.4 | 06.4 | 10.4	
02.4 | 08.4	2
03.4 | 07.4 | 09.4	
01.3	
05.3 | 10.3	
04.3 | 06.3	
02.3 | 08.3	0
03.3 | 07.3 | 09.3	
01.2	
04.2 | 05.2 | 10.2	
06.2	
02.2 | 08.2	-2
03.2 | 07.2 | 09.2	
	-4

Logits

Item Distriubtion

Data Science Practices (AFTER)

Person Distribution

	4
01.4 | 05.4	
04.4 | 06.4 | 10.4	
02.4 | 08.4	2
03.4 | 07.4 | 09.4	
01.3 | 05.3	
04.3 | 06.3 | 10.3	
02.3 | 08.3	0
03.3 | 07.3 | 09.3	
01.2	
05.2	
04.2 | 06.2 | 10.2	
02.2 | 08.2	-2
03.2 | 07.2 | 09.2	

Logits

Item Distriubtion
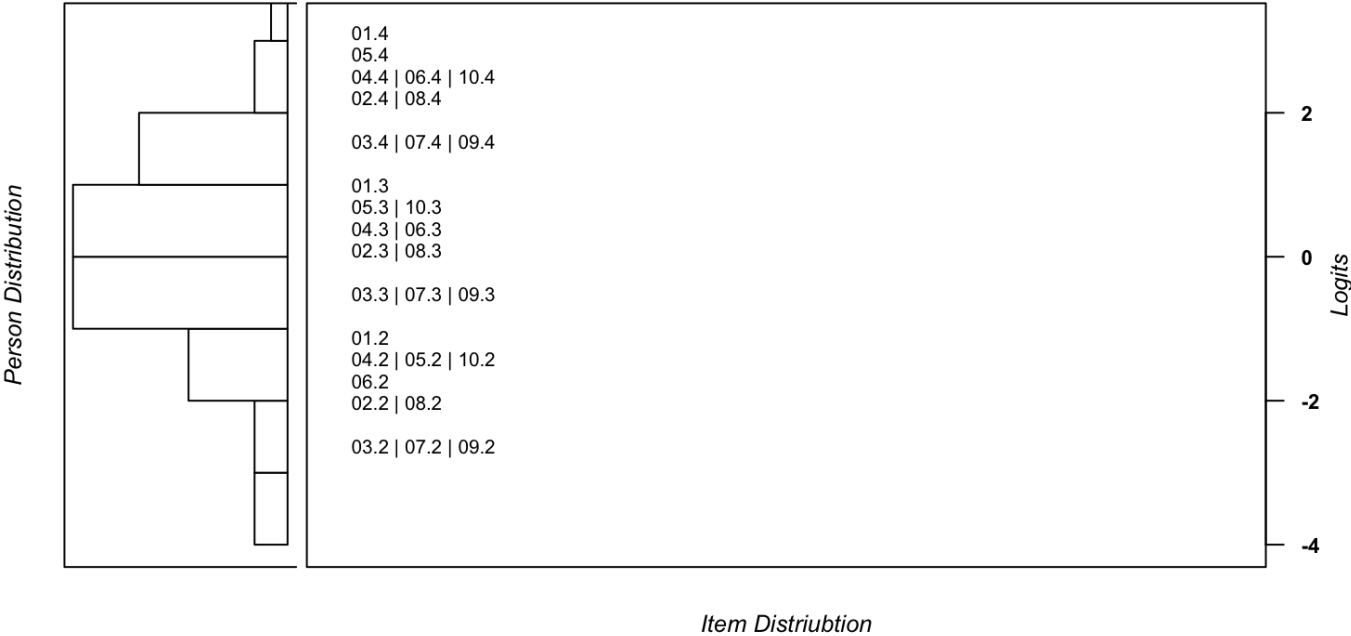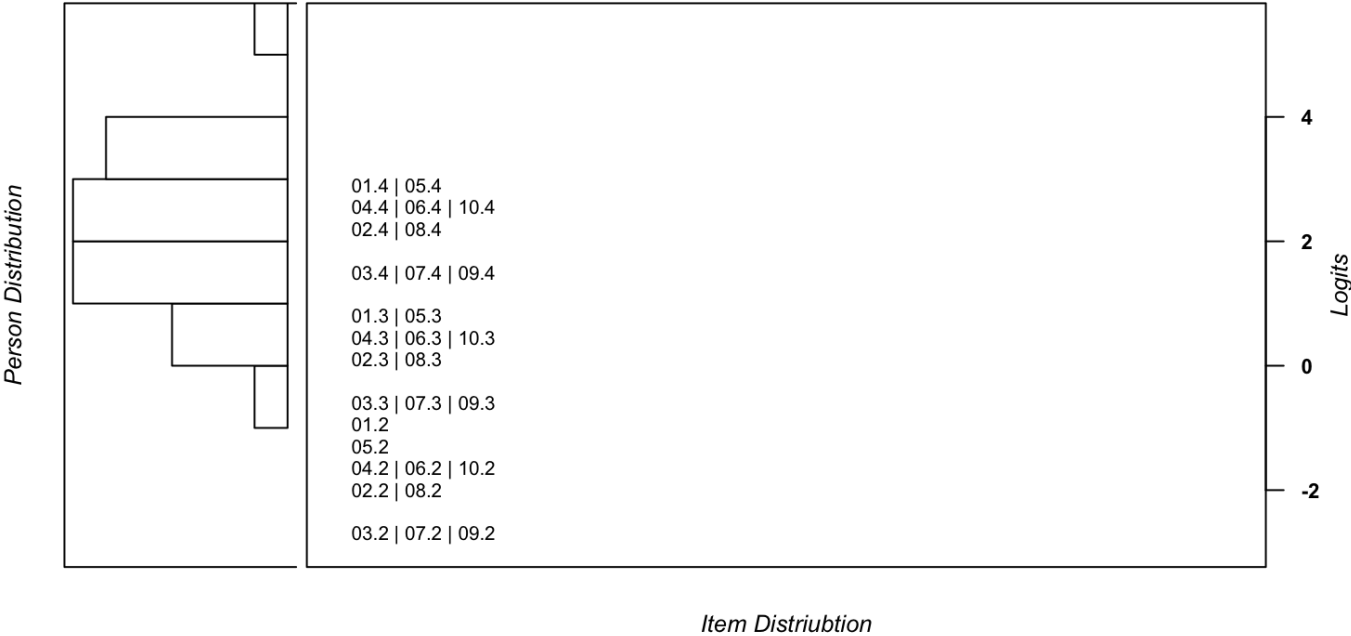
Science Identity

Science Identity (BEFORE)

Person Distribution

```
02.5
03.5 | 04.5
05.5
01.5 | 02.4
04.4
03.4 | 05.4
01.4


02.3
04.3
03.3 | 05.3
01.3

02.2
03.2 | 04.2
05.2
01.2
```

4

2

0   *Logits*

-2

-4

Item Distriubtion

Science Identity (AFTER)

Person Distribution

```
02.5
04.5
03.5 | 05.5
01.5
02.4
04.4
03.4 | 05.4
01.4


02.3
03.3 | 04.3
05.3
01.3

02.2
03.2 | 04.2
05.2
01.2
```

4

2

0   *Logits*

-2

-4

Item Distriubtion
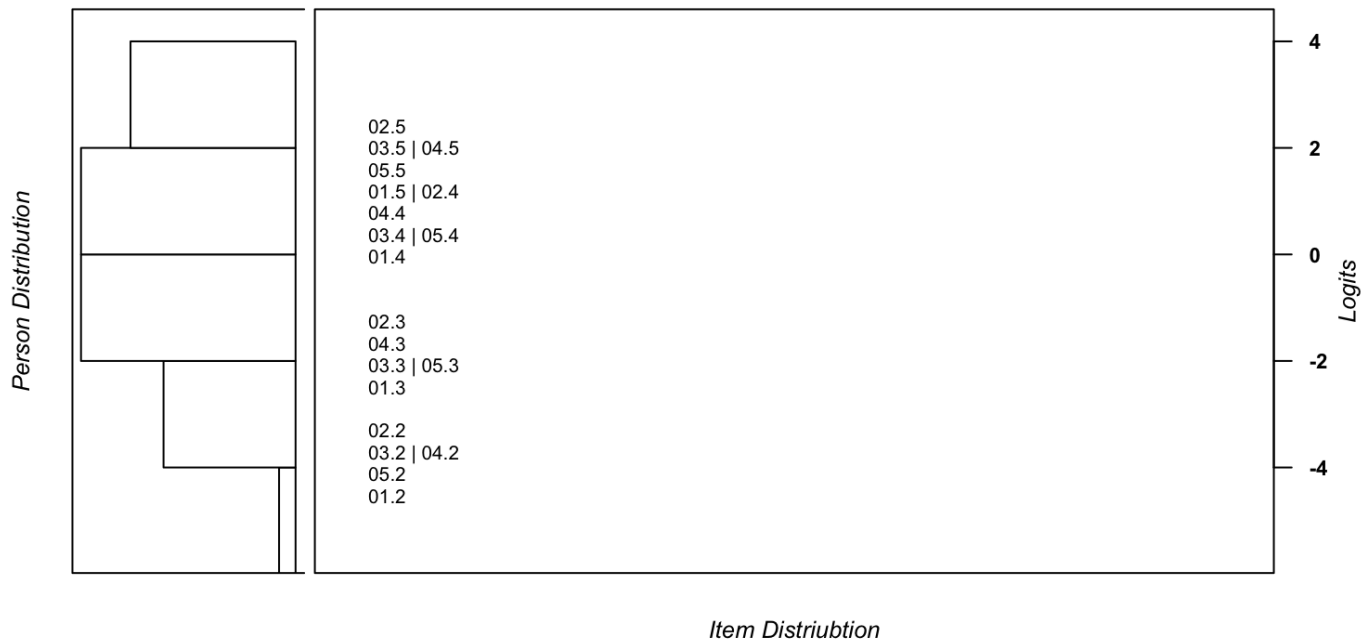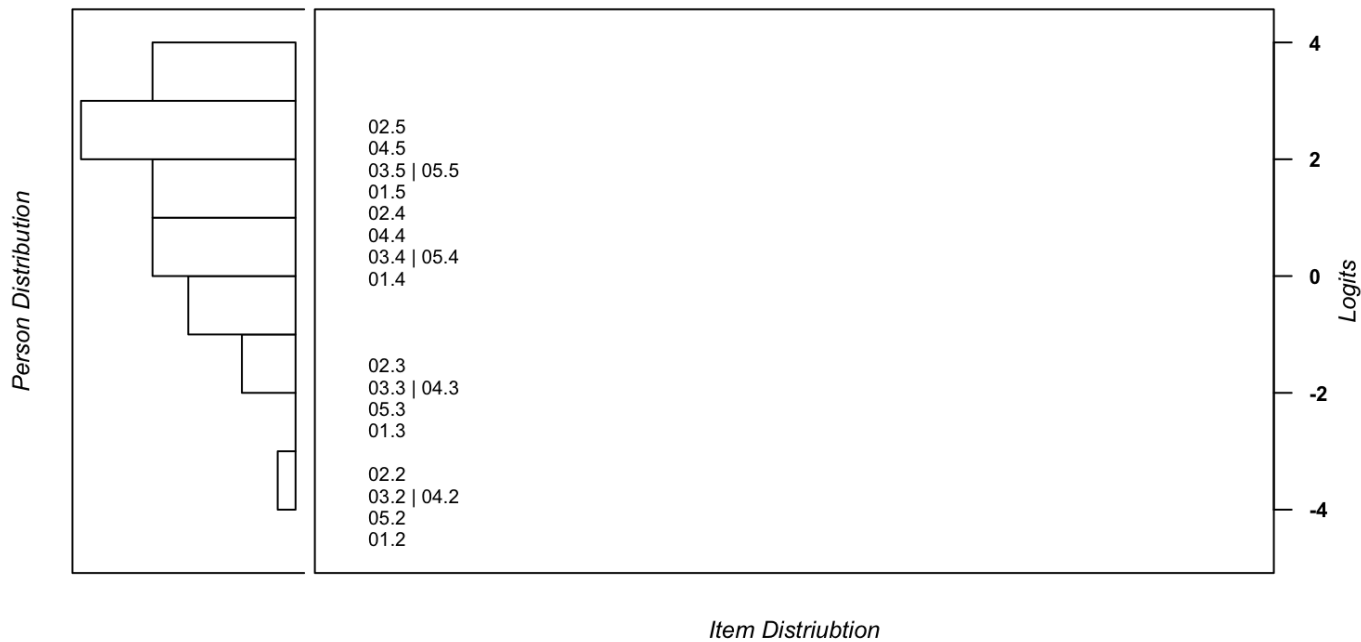
Analysis of Variance

```
##
## Call:
## lm(formula = Ability ~ Trial + Dimension + Cohort + Cohort *
##     Dimension, data = abil_trial_dimension_all6)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.228 -1.128  0.032  0.982  5.022
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                   2.75115    0.32182   8.549
## TrialBefore                                  -2.93001    0.16283 -17.994
## DimensionData Science Practices              -0.05436    0.44032  -0.123
## DimensionData Science Skills                  0.16079    0.44032   0.365
## DimensionPython Skills                        2.11823    0.44032   4.811
## DimensionScience Identity                    -0.12235    0.44032  -0.278
## CohortYear 2                                 -0.09295    0.46503  -0.200
## CohortYear 3                                 -0.13972    0.42259  -0.331
## DimensionData Science Practices:CohortYear 2 -0.24793    0.65766  -0.377
## DimensionData Science Skills:CohortYear 2     0.32802    0.66517   0.493
## DimensionPython Skills:CohortYear 2           0.72306    0.66517   1.087
## DimensionScience Identity:CohortYear 2       -0.59235    0.66517  -0.891
## DimensionData Science Practices:CohortYear 3 -0.12938    0.59763  -0.216
## DimensionData Science Skills:CohortYear 3     0.12335    0.60141   0.205
## DimensionPython Skills:CohortYear 3           0.70361    0.60141   1.170
## DimensionScience Identity:CohortYear 3       -0.62037    0.60141  -1.032
##                                              Pr(>|t|)
## (Intercept)                                   < 2e-16 ***
## TrialBefore                                   < 2e-16 ***
## DimensionData Science Practices                 0.902
## DimensionData Science Skills                    0.715
## DimensionPython Skills                        2.05e-06 ***
## DimensionScience Identity                       0.781
## CohortYear 2                                    0.842
## CohortYear 3                                    0.741
## DimensionData Science Practices:CohortYear 2    0.706
## DimensionData Science Skills:CohortYear 2       0.622
## DimensionPython Skills:CohortYear 2             0.278
## DimensionScience Identity:CohortYear 2          0.374
## DimensionData Science Practices:CohortYear 3    0.829
## DimensionData Science Skills:CohortYear 3       0.838
## DimensionPython Skills:CohortYear 3             0.243
## DimensionScience Identity:CohortYear 3          0.303
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.761 on 452 degrees of freedom
## Multiple R-squared:  0.5314, Adjusted R-squared:  0.5158
## F-statistic: 34.17 on 15 and 452 DF,  p-value: < 2.2e-16
```

```
## Analysis of Variance Table
##
## Response: Ability
##                   Df  Sum Sq Mean Sq  F value Pr(>F)
## Trial              1 1004.44 1004.44 323.7867 <2e-16 ***
## Dimension          4  563.71  140.93  45.4291 <2e-16 ***
## Cohort             2    1.34    0.67   0.2158 0.8060
## Dimension:Cohort   8   20.59    2.57   0.8295 0.5769
## Residuals        452 1402.18    3.10
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##                      eta.sq  eta.sq.part
## Trial            0.3356796615 0.4173656492
## Dimension        0.1883110526 0.2866604141
## Cohort           0.0004473933 0.0009538292
## Dimension:Cohort 0.0068797627 0.0144690270
```

t-testing & Cohen's *d*

Data Science Communication

```
##
##  Welch Two Sample t-test
##
## data:  comfortpost$Ability and comfortpre$Ability
## t = 10.27, df = 87.882, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.950954 2.887157
## sample estimates:
##    mean of x    mean of y
##  2.415192354 -0.003863267
```

```
## [1] 2.096379
```

Data Science Practices

```
##
##   Welch Two Sample t-test
##
## data:  confpost$Ability and confpre$Ability
## t = 7.5006, df = 92.239, p-value = 3.837e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.531676 2.634901
## sample estimates:
##    mean of x    mean of y
##  2.074584526 -0.008704105
```

```
## [1] 1.531062
```

Science Identity

```
##
##   Welch Two Sample t-test
##
## data:  idpost$Ability and idpre$Ability
## t = 3.3174, df = 82.505, p-value = 0.001353
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.6008267 2.4004040
## sample estimates:
##    mean of x    mean of y
##  1.43789668 -0.06271865
```

```
## [1] 0.6917187
```

Data Science Skills

```
##
##   Welch Two Sample t-test
##
## data:  techconfpost$Ability and techconfpre$Ability
## t = 10.45, df = 89.127, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.431845 3.573682
## sample estimates:
##    mean of x    mean of y
##  3.0032369769 0.0004732148
```
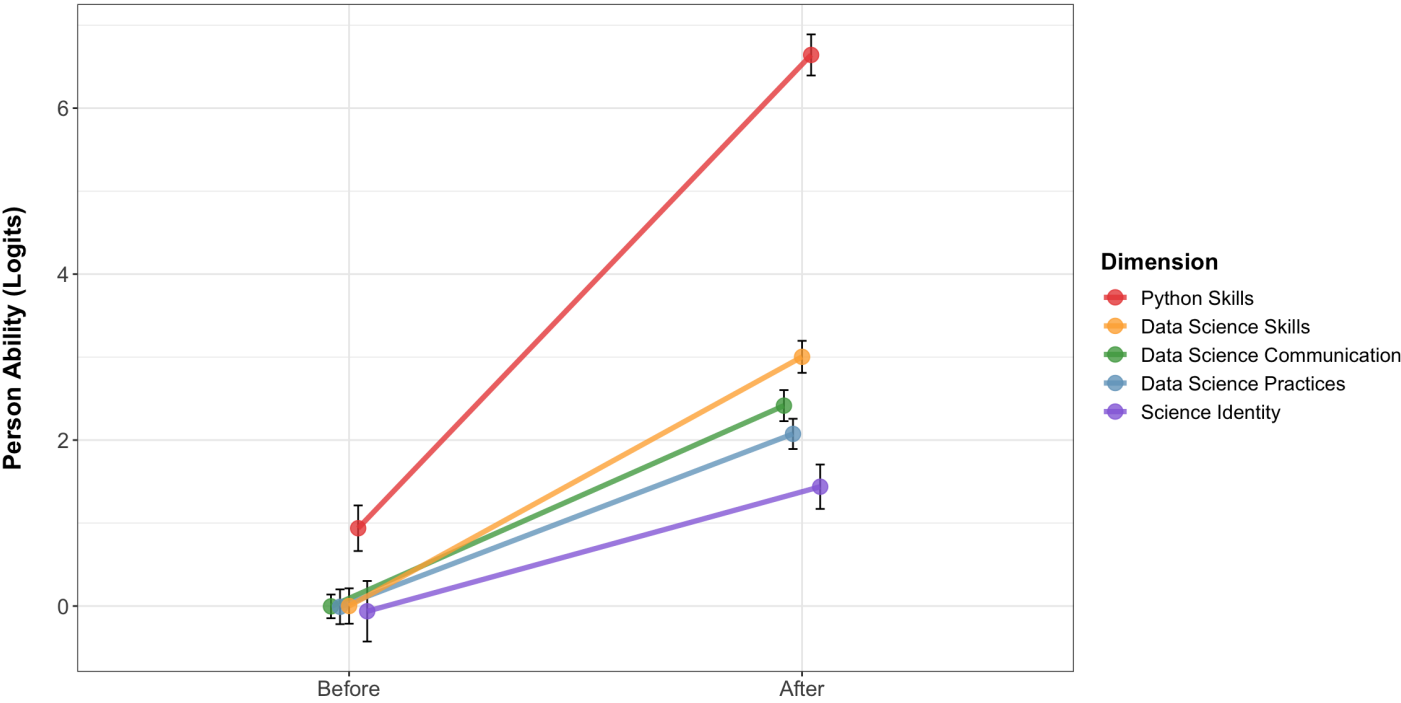
```
## [1] 2.179052
```

Python Skills

```
##
##   Welch Two Sample t-test
##
## data:  techcomfortpost$Ability and techcomfortpre$Ability
## t = 15.426, df = 89.051, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   4.968754 6.437964
## sample estimates:
## mean of x mean of y
##   6.641079  0.937720
```

```
## [1] 3.216642
```

Change in Ability by Dimension & Trial (Before/After)
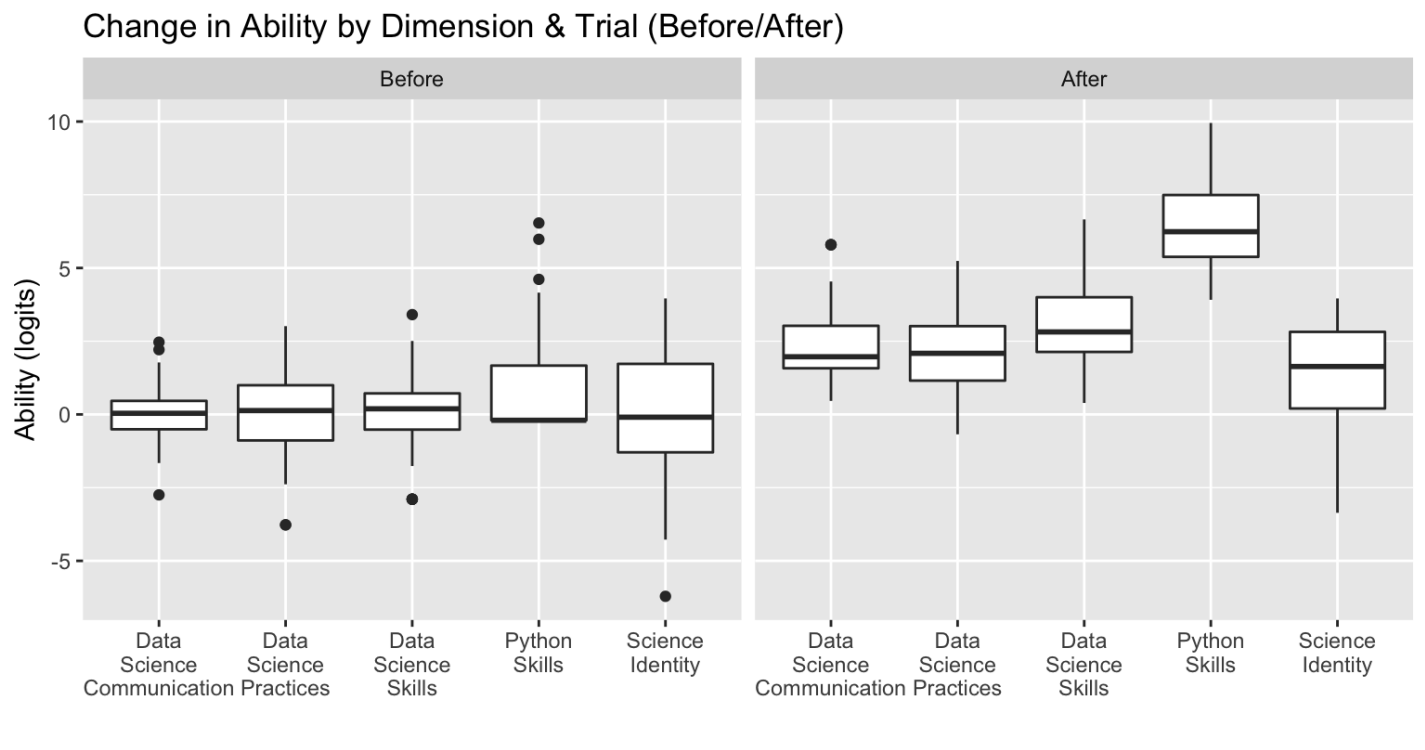
Growth across the five dimensions

Splitting by cohort