Report title

Subtitle

Report prepared for Black Saber Software by Eminence Analytics

2021-04-21

Contents


```
## Warning: package 'gtsummary' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.4
```

```
## Warning: package 'tibble' was built under R version 4.0.4
```

General comments (you can delete this section)

You can delete this section, and if you want to check what it said, just open a template from the package again. You don't have to use this particular template, but you DO need to write you report in RMarkdown and include a cover page.

The cover page must have:

- *A title and subtitle*
- *"Report prepared for Black Saber Software by" your company name*

- *Date (assessment submission date is fine)*

You can change the colour of this cover to any colour you would like by replacing 6C3082 in the YAML above (line 11) to another hex code. You could use this tool to help you: https://htmlcolorcodes.com/color-picker/

Executive summary

Guidelines for the executive summary:

- *No more than two pages*
- *Language is appropriate for a non-technical audience*
- *Bullet points are used where appropriate*
- *A small number of key visualizations and/or tables are included*
- *All three research questions are addressed*

Technical report

This part of the report is much more comprehensive than the executive summary. The audience is statistics/data-minded people, but you should NOT include code or unformatted R output here.

Introduction

Provide a brief introduction to your report and outline what the report will cover. This section is valuable for setting scope and expectations.

Research questions

Use bullet points to to describe the research questions you are going to address. Write in full sentences.

Informative title for section addressing a research question

For each research question, you will want to briefly describe any data manipulation, show some exploratory plots/summary tables, report on any methods you use (i.e. models you fit) and the conclusions you draw from these

```
## Joining, by = c("applicant_id", "team_applied_for", "cover_letter", "cv", "gpa", "gender", "extracur
```

```
## Joining, by = c("applicant_id", "team_applied_for", "cover_letter", "cv", "gpa", "gender", "extracur
```

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**Man**, N = 291	**Prefer not to say**, N = 11	**Woman**, N = 311	**p-value**
applicant_id	4,120 (2,705, 5,445)	3,950 (2,220, 5,240)	4,020 (2,435, 5,650)	0.8
team_applied_for				0.2
Data	125 (43%)	7 (64%)	124 (40%)	
Software	166 (57%)	4 (36%)	187 (60%)	
cover_letter	184 (63%)	7 (64%)	203 (65%)	0.9
cv	261 (90%)	7 (64%)	274 (88%)	0.045
gpa	2.70 (2.20, 3.25)	2.40 (2.10, 2.65)	2.70 (2.20, 3.30)	0.5
extracurriculars				0.5
0	26 (8.9%)	1 (9.1%)	26 (8.4%)	
1	178 (61%)	9 (82%)	184 (59%)	
2	87 (30%)	1 (9.1%)	101 (32%)	
work_experience				0.8
0	43 (15%)	2 (18%)	50 (16%)	
1	218 (75%)	9 (82%)	223 (72%)	
2	30 (10%)	0 (0%)	38 (12%)	
result				0.3
proceed	145 (50%)	3 (27%)	152 (49%)	
refused	146 (50%)	8 (73%)	159 (51%)	

```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	**Man**, N = 145	**Prefer not to say**, N = 3	**Woman**, N = 152
applicant_id	4,237.17 (1,836.63)	3,900.00 (1,755.53)	4,046.45 (1,766.42)
team_applied_for			
Data	67 / 145 (46%)	2 / 3 (67%)	63 / 152 (41%)
Software	78 / 145 (54%)	1 / 3 (33%)	89 / 152 (59%)
cover_letter	145 / 145 (100%)	3 / 3 (100%)	152 / 152 (100%)
cv	145 / 145 (100%)	3 / 3 (100%)	152 / 152 (100%)
gpa	3.08 (0.54)	2.67 (0.46)	3.14 (0.54)
extracurriculars			
1	90 / 145 (62%)	2 / 3 (67%)	87 / 152 (57%)
2	55 / 145 (38%)	1 / 3 (33%)	65 / 152 (43%)
work_experience			
0	3 / 145 (2.1%)	0 / 3 (0%)	6 / 152 (3.9%)
1	119 / 145 (82%)	3 / 3 (100%)	114 / 152 (75%)
2	23 / 145 (16%)	0 / 3 (0%)	32 / 152 (21%)
technical_skills	46.38 (22.16)	66.00 (13.53)	47.98 (21.45)
writing_skills	41.78 (18.05)	42.33 (12.06)	46.19 (19.35)
leadership_presence	4.92 (2.53)	4.33 (4.16)	4.07 (2.05)
speaking_skills	5.08 (2.39)	6.67 (2.08)	3.31 (2.05)
result			
proceed	15 / 145 (10%)	0 / 3 (0%)	7 / 152 (4.6%)
refused	130 / 145 (90%)	3 / 3 (100%)	145 / 152 (95%)

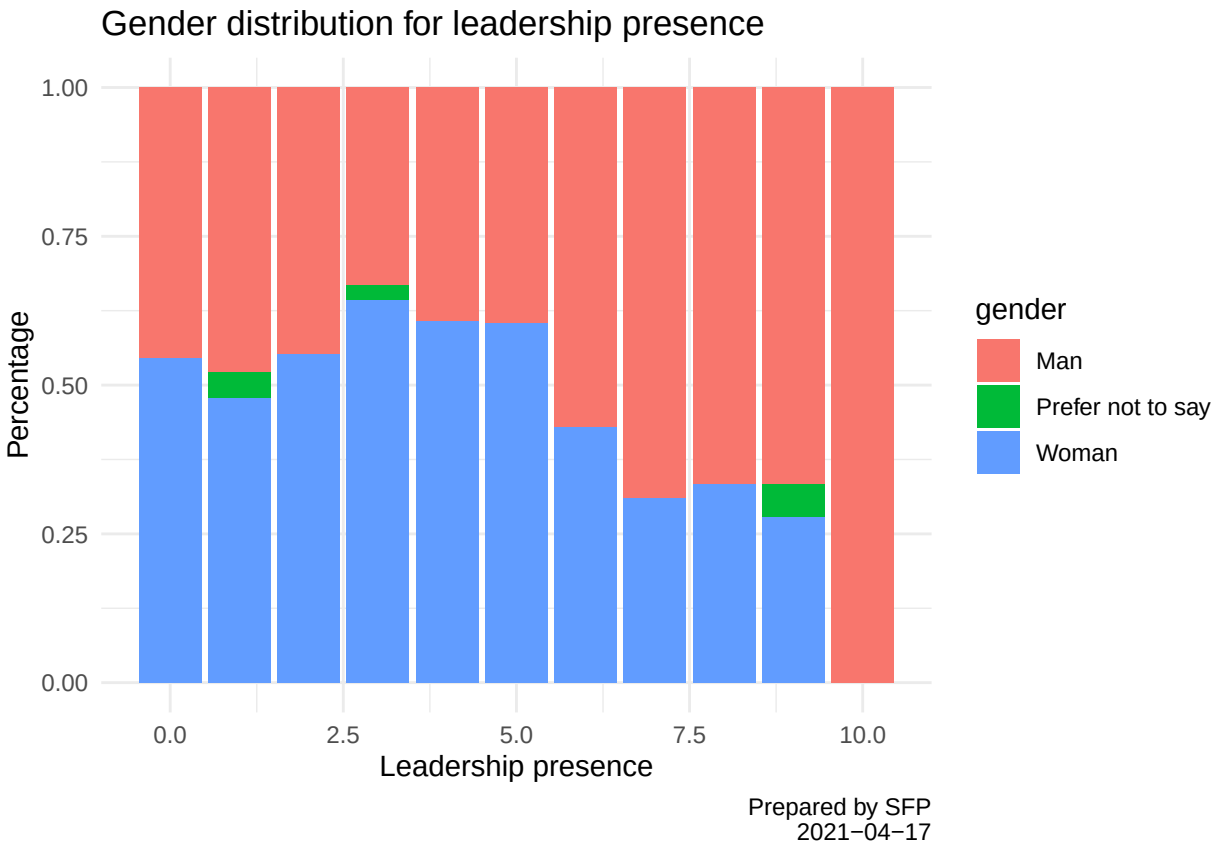
```
## Table printed with `knitr::kable()`, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

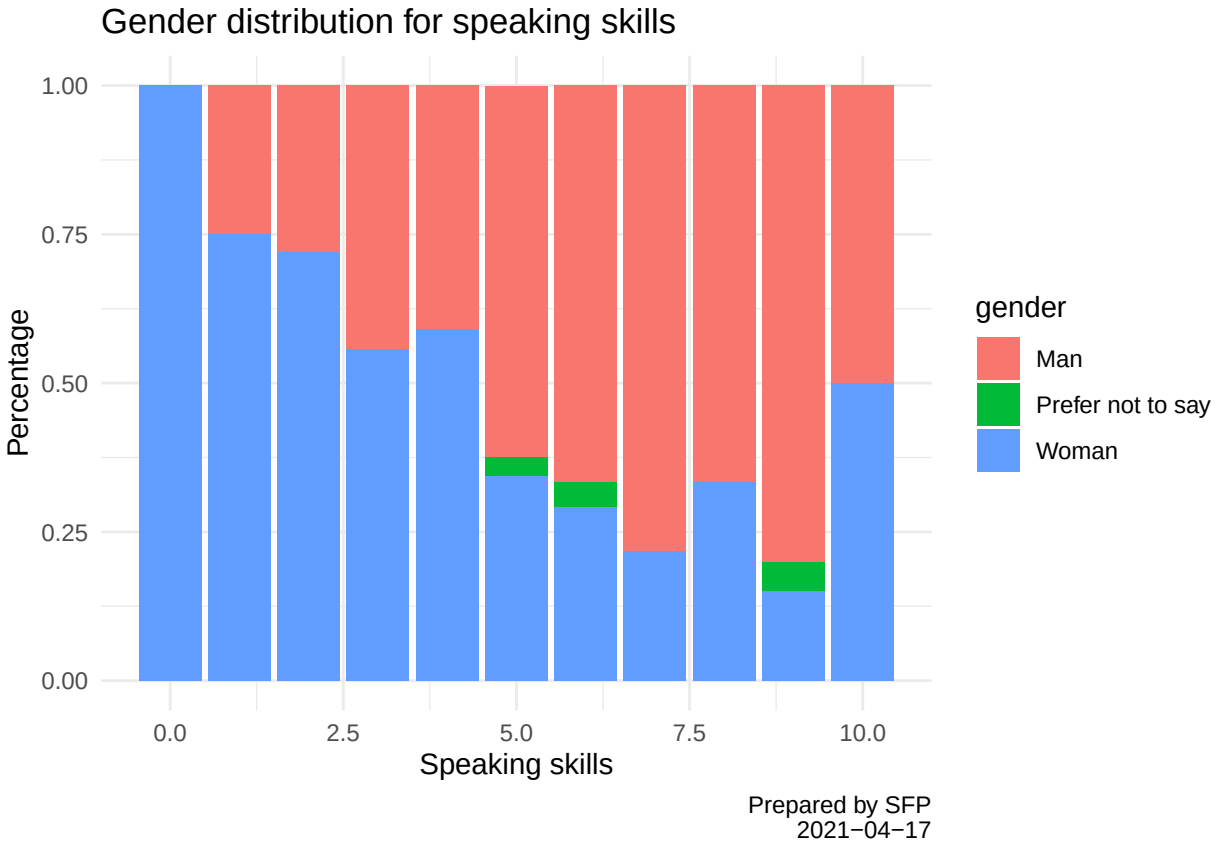
Characteristic	**Man**, N = 291	**Prefer not to say**, N = 11	**Woman**, N = 311
applicant_id	4,101.07 (1,744.82)	3,766.36 (1,949.83)	4,051.67 (1,793.59)
team_applied_for			
Data	125 / 291 (43%)	7 / 11 (64%)	124 / 311 (40%)
Software	166 / 291 (57%)	4 / 11 (36%)	187 / 311 (60%)
cover_letter	184 / 291 (63%)	7 / 11 (64%)	203 / 311 (65%)
cv	261 / 291 (90%)	7 / 11 (64%)	274 / 311 (88%)
gpa	2.71 (0.67)	2.50 (0.58)	2.72 (0.67)
extracurriculars			
0	26 / 291 (8.9%)	1 / 11 (9.1%)	26 / 311 (8.4%)
1	178 / 291 (61%)	9 / 11 (82%)	184 / 311 (59%)
2	87 / 291 (30%)	1 / 11 (9.1%)	101 / 311 (32%)
work_experience			
0	43 / 291 (15%)	2 / 11 (18%)	50 / 311 (16%)
1	218 / 291 (75%)	9 / 11 (82%)	223 / 311 (72%)
2	30 / 291 (10%)	0 / 11 (0%)	38 / 311 (12%)
technical_skills	46.38 (22.16)	66.00 (13.53)	47.98 (21.45)
(Missing)	146	8	159
writing_skills	41.78 (18.05)	42.33 (12.06)	46.19 (19.35)
(Missing)	146	8	159

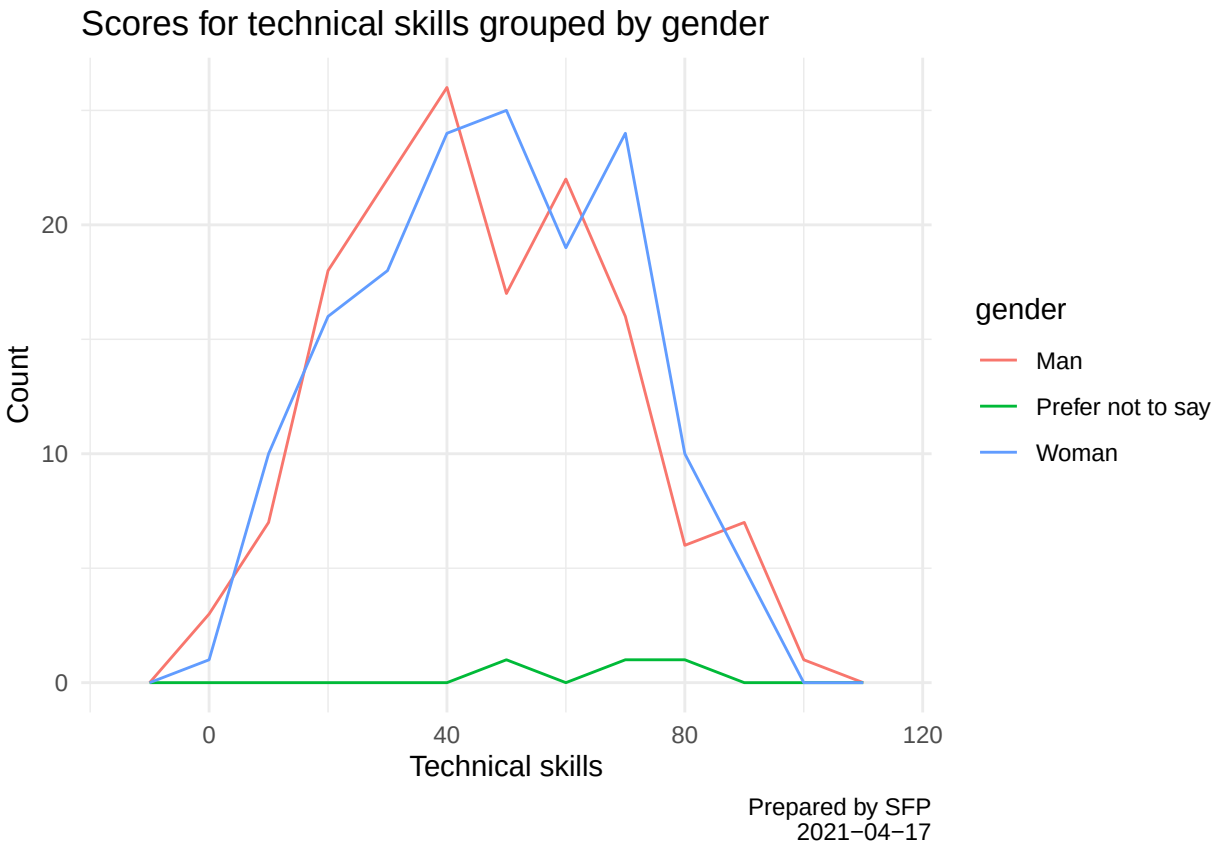
Characteristic	**Man**, N = 291	**Prefer not to say**, N = 11	**Woman**, N = 311
leadership_presence	4.92 (2.53)	4.33 (4.16)	4.07 (2.05)
(Missing)	146	8	159
speaking_skills	5.08 (2.39)	6.67 (2.08)	3.31 (2.05)
(Missing)	146	8	159
interviewer_rating_1	75.53 (5.84)	NA (NA)	74.43 (7.25)
(Missing)	276	11	304
interviewer_rating_2	75.33 (9.70)	NA (NA)	75.00 (7.92)
(Missing)	276	11	304
final_result			
Refused at phase 1	146 / 291 (50%)	8 / 11 (73%)	159 / 311 (51%)
Refused at phase 2	130 / 291 (45%)	3 / 11 (27%)	145 / 311 (47%)
Refused at phase 3	7 / 291 (2.4%)	0 / 11 (0%)	5 / 311 (1.6%)
Hired	8 / 291 (2.7%)	0 / 11 (0%)	2 / 311 (0.6%)
result			
Hired	8 / 291 (2.7%)	0 / 11 (0%)	2 / 311 (0.6%)
no	283 / 291 (97%)	11 / 11 (100%)	309 / 311 (99%)

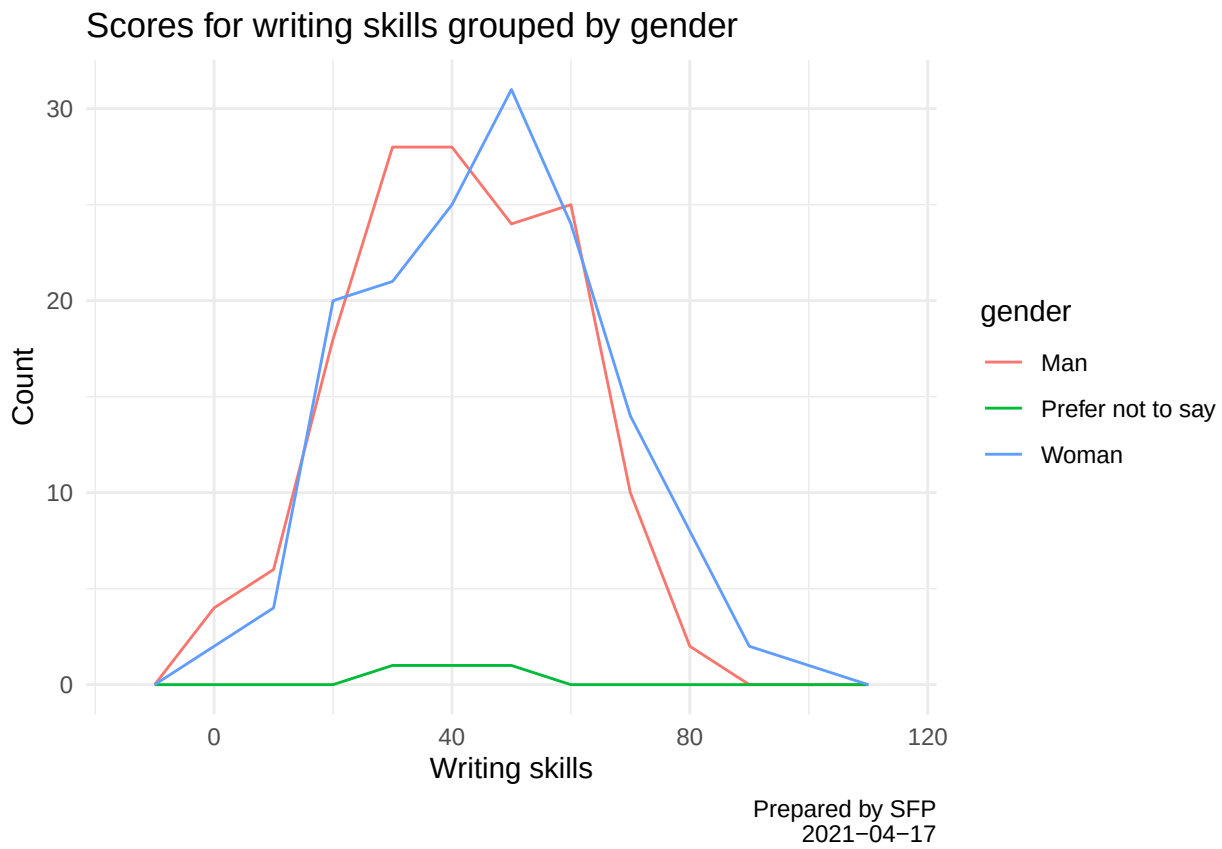

Interview rating 1 grouped by gender
gender
Man
Woman
60
70
80
Interview rating 1
Prepared by SFP
2021−04−17


Interview rating 2 grouped by gender
gender
Man
Woman
60
70
80
90
Interview rating 2
Prepared by SFP
2021−04−17


Gender distribution for leadership presence
1.00
0.75
0.50
0.25
0.00
Percentage
0.0
2.5
5.0
7.5
10.0
Leadership presence
Prepared by SFP
2021−04−17
gender
Man
Prefer not to say
Woman


Gender distribution for speaking skills
1.00
0.75
0.50
0.25
0.00
Percentage
0.0
2.5
5.0
7.5
10.0
Speaking skills
Prepared by SFP
2021−04−17
gender
Man
Prefer not to say
Woman


Scores for technical skills grouped by gender
Count
20
10
0
0
40
80
120
Technical skills
Prepared by SFP
2021−04−17
gender
Man
Prefer not to say
Woman




```
##
## Call:
## glm(formula = as.factor(result) ~ gpa + as.factor(gender) + extracurriculars +
##     work_experience, family = binomial("logit"), data = phase1)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0573  -0.7890   0.2768   0.7159   2.4499
##
## Coefficients:
##                                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)                        6.40374    0.53772  11.909   <2e-16 ***
## gpa                               -1.97580    0.21076  -9.375   <2e-16 ***
## as.factor(gender)Prefer not to say  0.57889    0.79473   0.728   0.4664
## as.factor(gender)Woman              0.06458    0.20361   0.317   0.7511
## extracurriculars                   -0.26442    0.19616  -1.348   0.1777
## work_experience                    -0.75728    0.25693  -2.947   0.0032 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 849.52  on 612  degrees of freedom
## Residual deviance: 600.83  on 607  degrees of freedom
## AIC: 612.83
##
```

```
## Number of Fisher Scoring iterations: 4

## Table printed with `knitr::kable()`, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	OR	95% CI	p-value
gpa	0.14	0.09, 0.21	<0.001
as.factor(gender)			
Man			
Prefer not to say	1.78	0.42, 10.1	0.5
Woman	1.07	0.72, 1.59	0.8
extracurriculars	0.77	0.52, 1.13	0.2
work_experience	0.47	0.28, 0.76	0.003

```
##
## Call:
## glm(formula = as.factor(result) ~ gpa + as.factor(gender) + extracurriculars +
##     work_experience + leadership_presence + speaking_skills +
##     technical_skills + writing_skills, family = binomial("logit"),
##     data = phase2)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.98920   0.00767   0.03916   0.15889   2.01541
##
## Coefficients:
##                                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)                        19.90091    3.91877   5.078 3.81e-07 ***
## gpa                                 0.77031    0.76231   1.010  0.31226
## as.factor(gender)Prefer not to say 16.61510 2031.56549   0.008  0.99347
## as.factor(gender)Woman              0.52264    0.75916   0.688  0.49118
## extracurriculars                    0.35768    0.70999   0.504  0.61442
## work_experience                    -0.01173    0.77207  -0.015  0.98788
## leadership_presence                -0.97439    0.21877  -4.454 8.43e-06 ***
## speaking_skills                    -0.76314    0.17979  -4.245 2.19e-05 ***
## technical_skills                   -0.09402    0.02469  -3.808  0.00014 ***
## writing_skills                     -0.10349    0.02639  -3.921 8.81e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 157.31  on 299  degrees of freedom
## Residual deviance:  67.31  on 290  degrees of freedom
## AIC: 87.31
##
## Number of Fisher Scoring iterations: 16

## Table printed with `knitr::kable()`, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

Characteristic	OR	95% CI	p-value
gpa	2.16	0.51, 10.7	0.3
as.factor(gender)			
Man			
Prefer not to say	16,437,923	0.00, NA	>0.9
Woman	1.69	0.38, 7.87	0.5
extracurriculars	1.43	0.36, 6.15	0.6
work_experience	0.99	0.22, 4.85	>0.9
leadership_presence	0.38	0.23, 0.55	<0.001
speaking_skills	0.47	0.31, 0.64	<0.001
technical_skills	0.91	0.86, 0.95	<0.001
writing_skills	0.90	0.85, 0.94	<0.001

```
##
## Call:
## glm(formula = as.factor(result) ~ gpa + as.factor(gender) + extracurriculars +
##     work_experience + leadership_presence + speaking_skills +
##     technical_skills + writing_skills + interviewer_rating_2,
##     family = binomial("logit"), data = phase3)
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -1.163e-05  -2.110e-08   2.110e-08   2.110e-08   1.156e-05
##
## Coefficients:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)              7.686e+02  1.227e+06   0.001        1
## gpa                     -2.889e+01  3.420e+05   0.000        1
## as.factor(gender)Woman   3.489e+01  2.582e+05   0.000        1
## extracurriculars        -7.599e+00  2.973e+05   0.000        1
## work_experience         -4.766e+01  1.329e+05   0.000        1
## leadership_presence     -1.761e+01  7.961e+04   0.000        1
## speaking_skills         -1.526e+01  7.378e+04   0.000        1
## technical_skills        -1.379e-01  1.183e+04   0.000        1
## writing_skills           2.135e+00  1.756e+04   0.000        1
## interviewer_rating_2    -6.461e+00  1.326e+04   0.000        1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3.0316e+01  on 21  degrees of freedom
## Residual deviance: 7.6669e-10  on 12  degrees of freedom
## AIC: 20
##
## Number of Fisher Scoring iterations: 25

## Table printed with `knitr::kable()`, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

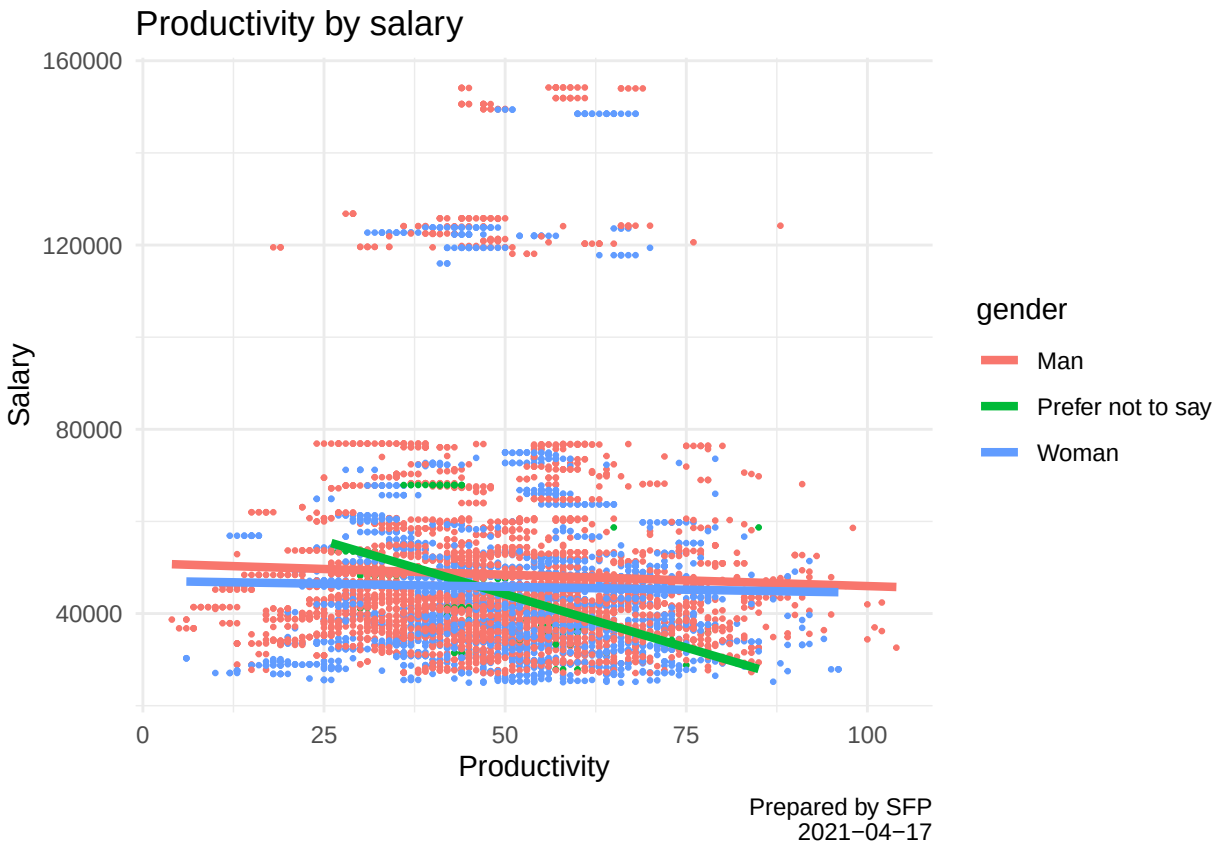
Characteristic	OR	95% CI	p-value
gpa	0.00	0.00, Inf	>0.9
as.factor(gender)			
Man			
Woman	1,420,556,165,452,708		>0.9
extracurriculars	0.00	0.00, Inf	>0.9
work_experience	0.00	0.00, Inf	>0.9
leadership_presence	0.00		>0.9
speaking_skills	0.00	0.00, Inf	>0.9
technical_skills	0.87	0.00, Inf	>0.9
writing_skills	8.46	0.00, Inf	>0.9
interviewer_rating_2	0.00	0.00, Inf	>0.9

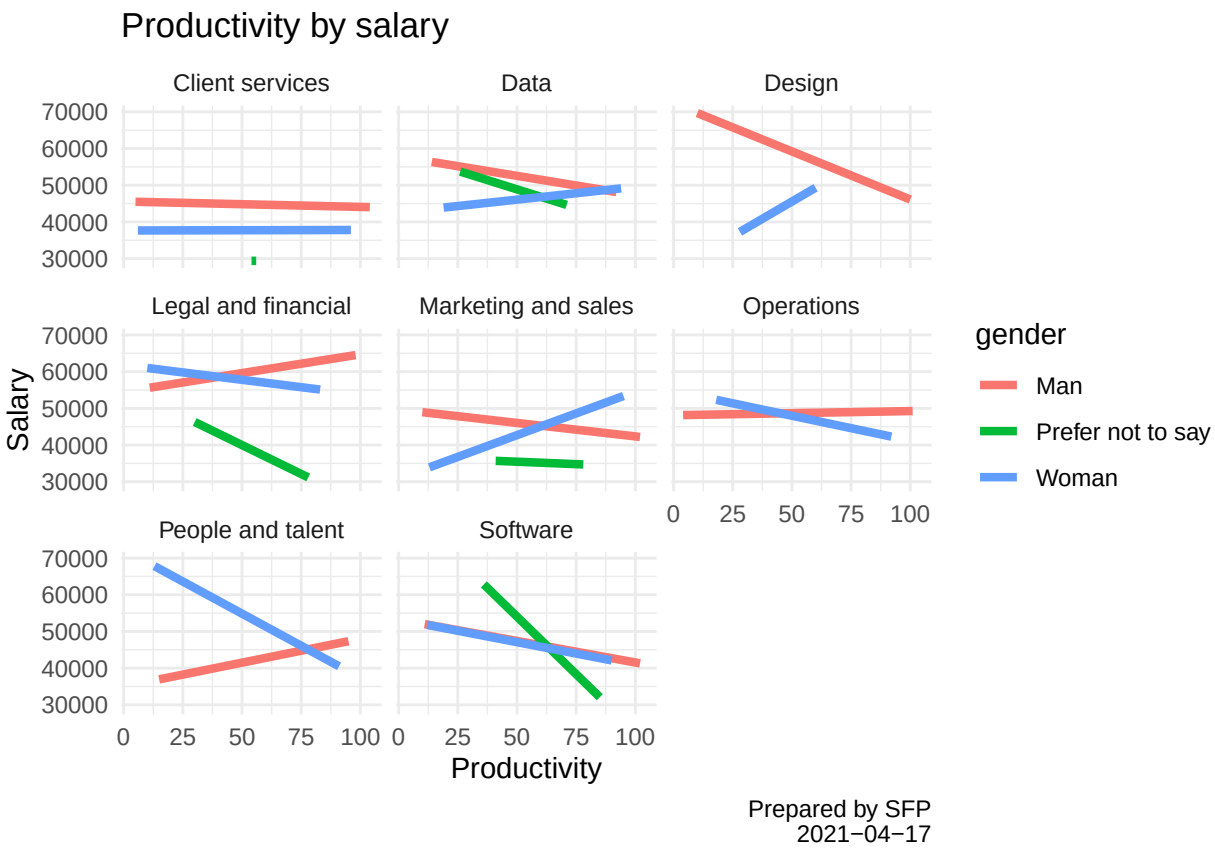
```
##
## Call:
## glm(formula = as.factor(result) ~ gpa + as.factor(gender) + extracurriculars +
##     work_experience + leadership_presence + speaking_skills +
##     technical_skills + writing_skills + interviewer_rating_1 +
##     interviewer_rating_2, family = binomial("logit"), data = all_phases)
##
## Deviance Residuals:
##        Min          1Q      Median          3Q         Max
## -1.045e-05  -1.739e-06   2.110e-08   2.928e-06   9.548e-06
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)             5.305e+02  1.288e+06       0        1
## gpa                     3.411e+00  2.778e+05       0        1
## as.factor(gender)Woman  5.817e+01  3.177e+05       0        1
## extracurriculars       -8.318e+00  1.985e+05       0        1
## work_experience         9.562e+00  2.205e+05       0        1
## leadership_presence     1.636e+02  9.451e+05       0        1
## speaking_skills         1.658e+02  8.931e+05       0        1
## technical_skills        1.695e+01  9.460e+04       0        1
## writing_skills          1.745e+01  9.229e+04       0        1
## interviewer_rating_1   -6.058e+01  3.221e+05       0        1
## interviewer_rating_2   -3.157e+00  1.543e+04       0        1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3.0316e+01  on 21  degrees of freedom
## Residual deviance: 5.3722e-10  on 11  degrees of freedom
##   (591 observations deleted due to missingness)
## AIC: 22
##
## Number of Fisher Scoring iterations: 25

## ! `broom::tidy()` failed to tidy the model.

## x need at least two non-NA values to interpolateapprox(sp$y, sp$x, xout = cutoff)

## v `tidy_parameters()` used instead.
```

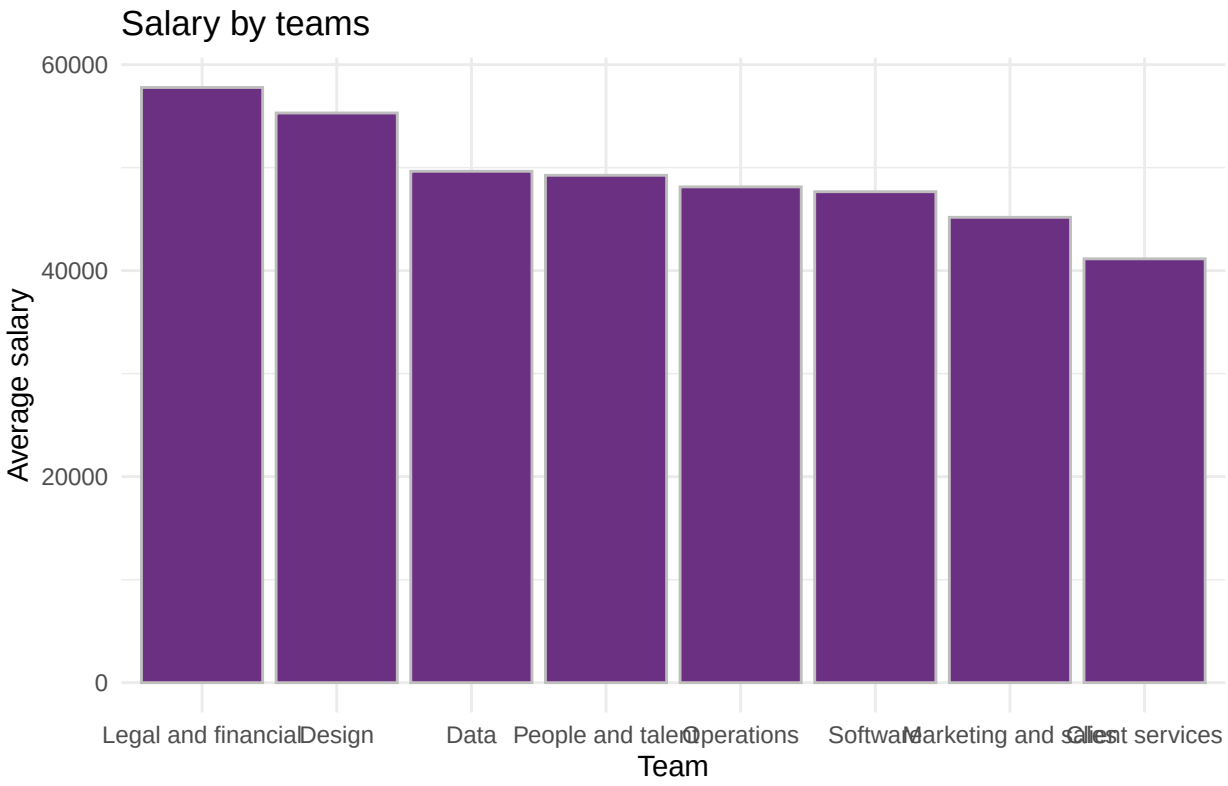
```
## i Add `tidy_fun = tidy_parameters` to quiet these messages.

## Table printed with `knitr::kable()`, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in code chunk header.
```

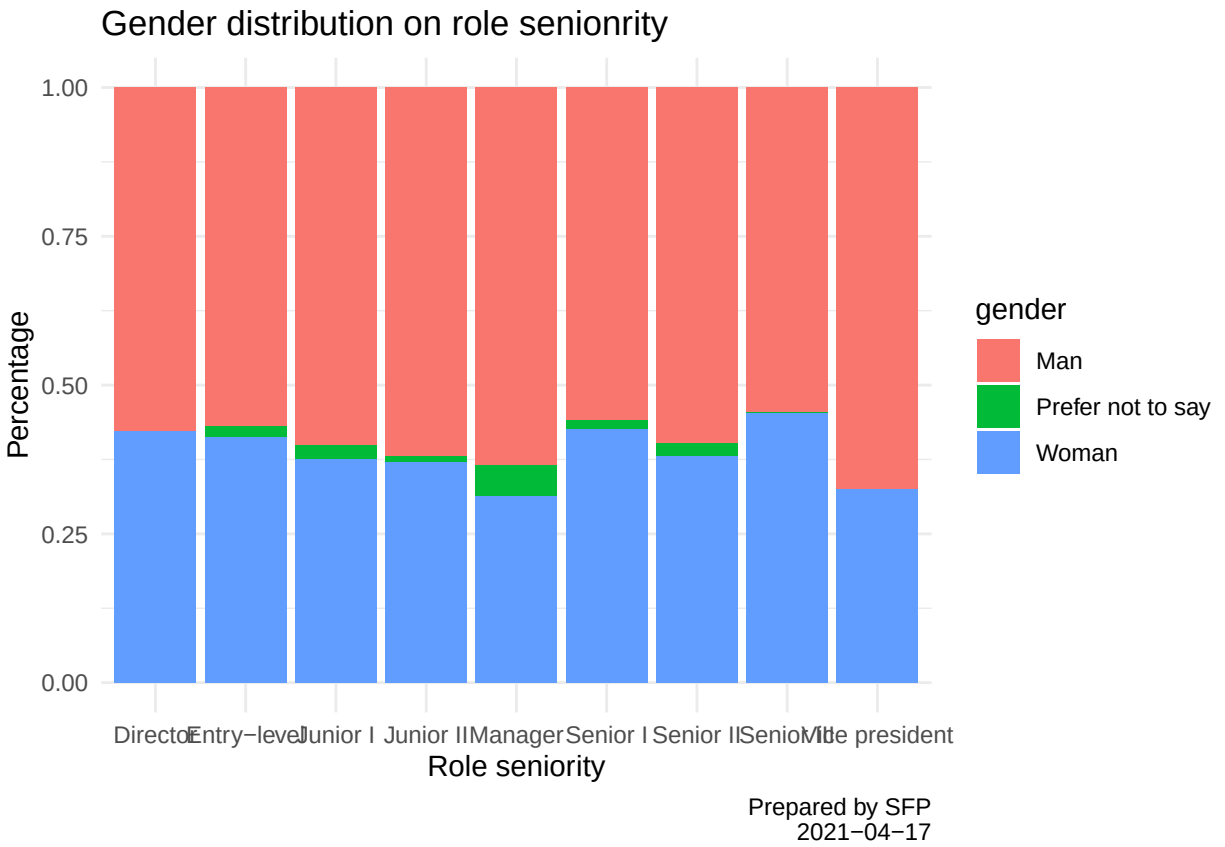
Characteristic	OR	95% CI	p-value
gpa	30.3	0.00, Inf	>0.9
as.factor(gender)			
Man			
Woman	18,286,077,681,272,950,710,462,086	0.00, Inf	>0.9
extracurriculars	0.00	0.00, Inf	>0.9
work_experience	14,215	0.00, Inf	>0.9
leadership_present	17,728,837,893,962,624,784,282,020,882,460,648,682,424,620,240,840,686,466,204,049	0.00, Inf	>0.9
speaking_skills	1,007,767,241,528,988,120,404,248,860,246,486,402,844,202,602,248,420,264,446,964,420,428,406,240	0.00, Inf	>0.9
technical_skills	22,934,249	0.00, Inf	>0.9
writing_skills	38,055,421	0.00, Inf	>0.9
interviewer_rating_1	0.01	0.00, Inf	>0.9
interviewer_rating_2	0.42	0.00, Inf	>0.9

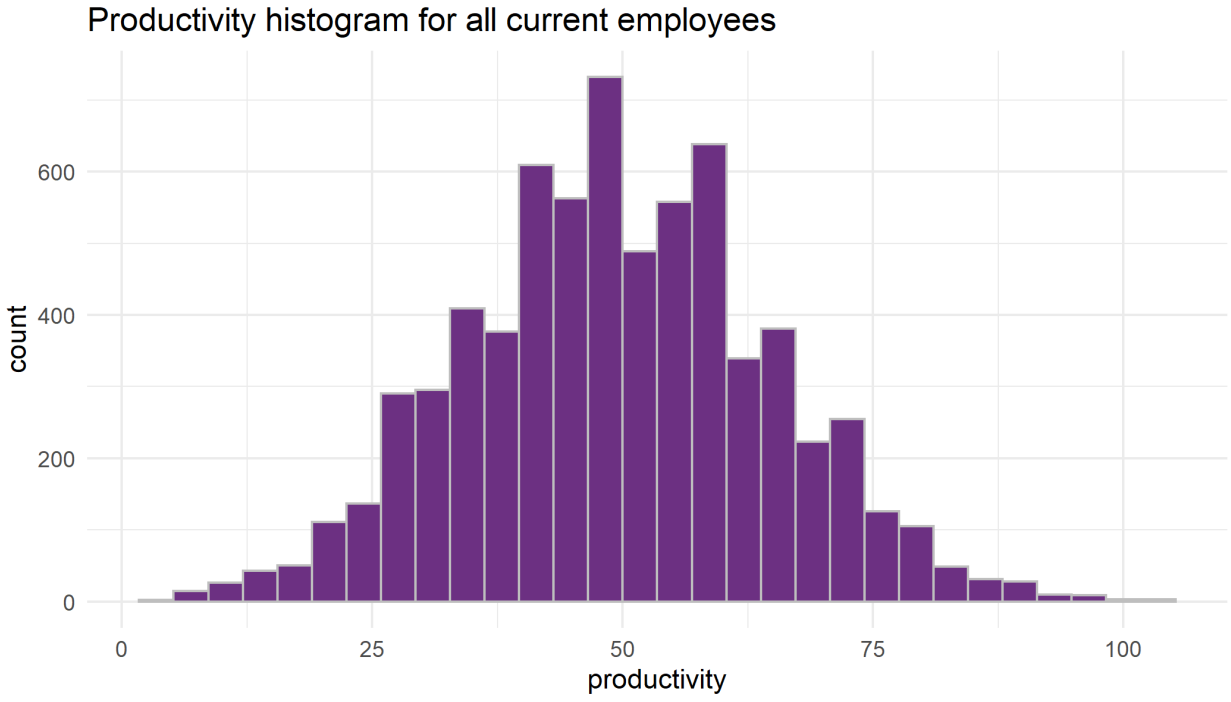
Characteristic	**Man**, N = 340	**Prefer not to say**, N = 10	**Woman**, N = 257
role_seniority			
Director	12 (3.5%)	0 (0%)	7 (2.7%)
Entry-level	79 (23%)	3 (30%)	96 (37%)
Junior I	69 (20%)	2 (20%)	53 (21%)
Junior II	53 (16%)	2 (20%)	17 (6.6%)
Manager	19 (5.6%)	1 (10%)	8 (3.1%)
Senior I	42 (12%)	1 (10%)	29 (11%)
Senior II	28 (8.2%)	1 (10%)	21 (8.2%)
Senior III	32 (9.4%)	0 (0%)	24 (9.3%)
Vice president	6 (1.8%)	0 (0%)	2 (0.8%)
productivity	50 (42, 60)	46 (42, 58)	54 (43, 64)
salary	41,150 (35,600, 50,125)	39,050 (34,100, 46,475)	34,900 (30,600, 44,600)


Productivity by salary
160000
120000
80000
40000
Salary
0
25
50
75
100
Productivity
Prepared by SFP
2021−04−17
gender
Man
Prefer not to say
Woman


Productivity by salary
Client services
Data
Design
Legal and financial
Marketing and sales
Operations
People and talent
Software
Salary
Productivity
70000
60000
50000
40000
30000
0
25
50
75
100
gender
Man
Prefer not to say
Woman
Prepared by SFP
2021−04−17


Salary by teams
60000
40000
20000
0
Average salary
Legal and financial
Design
Data
People and talent
Operations
Software
Marketing and sales
Client services
Team
Prepared by SFP
2021−04−17


Gender distribution on role seniorrity
1.00
0.75
0.50
0.25
0.00
Percentage
Director
Entry−level
Junior I
Junior II
Manager
Senior I
Senior II
Senior III
Vice president
Role seniority
gender
Man
Prefer not to say
Woman
Prepared by SFP
2021−04−17


Productivity histogram for all current employees
600
400
200
0
count
0
25
50
75
100
productivity

Informative title for section addressing a research question

Informative title for section addressing a research question

Discussion

In this section you will summarize your findings across all the research questions and discuss the strengths and limitations of your work. It doesn't have to be long, but keep in mind that often people will just skim the intro and the discussion of a document like this, so make sure it is useful as a semi-standalone section (doesn't have to be completely standalone like the executive summary).

Strengths and limitations

Consultant information

Consultant profiles

Complete this section with a brief bio for each member of your group. If you are completing the project individually, you only need to complete one for yourself. In that case, change the title of this section to 'Consultant profile' instead. Examples below. This section is only marked for completeness, clarity and professionalism, not 'truth' so you can write it as if we're a few years in the future. Put your current degree in as completed and/or add your first choice grad school program, whatever you like. What skills related skills would you most like to highlight? What job title do you want?

Lucas Xian. Lucas is a senior consultant with SFP. He specializes in data analysis and visualization. Lucas earned his Honours Bachelor of Science, Specialist in Statistics, from the University of Toronto in 2023.

Dandan Zhang. Dandan is a junior consultant with SFP. She earned her Bachelor of Science, majoring in Mathematics and Statistics, from the University of Toronto in 2021.She specializes in data visualization.

Code of ethical conduct

This section should be fairly short, no more than half a page. Assume a general audience, much like your executive summary.

- *Make at least three relevant statements about your company's approach to ethical statistical consulting. These should be appropriately in line with professional conduct advice like the (Statistical Society of Canada Code of Conduct)[https://ssc.ca/sites/default/files/data/Members/public/Accreditation/ethics_e.pdf] or the (Ethical Guidelines for Statistical Practice from the American Statistical Society)[https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx]. For example, "the customer is always right" ISN'T the type of thing an ethical statistical consultant would include.*
- *Be very careful not to just copy and paste from these other documents! Put things in your own words.*

Final advice: KNIT EARLY AND OFTEN!