

---

# Report title

Subtitle

Report prepared for Black Saber Software by Eminence  
Analytics

2021-04-21

## Contents

<b>General comments (you can delete this section)</b>	<b>2</b>
<b>Executive summary</b>	<b>3</b>
<b>Technical report</b>	<b>4</b>
Introduction . . . . .	4
Informative title for section addressing a research question . . . . .	4
Informative title for section addressing a research question . . . . .	5
Informative title for section addressing a research question . . . . .	5
Import Data . . . . .	5
Data Cleaning . . . . .	5
Data Visualization . . . . .	5
Discussion . . . . .	15
<b>Consultant information</b>	<b>16</b>
Consultant profiles . . . . .	16
Code of ethical conduct . . . . .	16

## 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.*

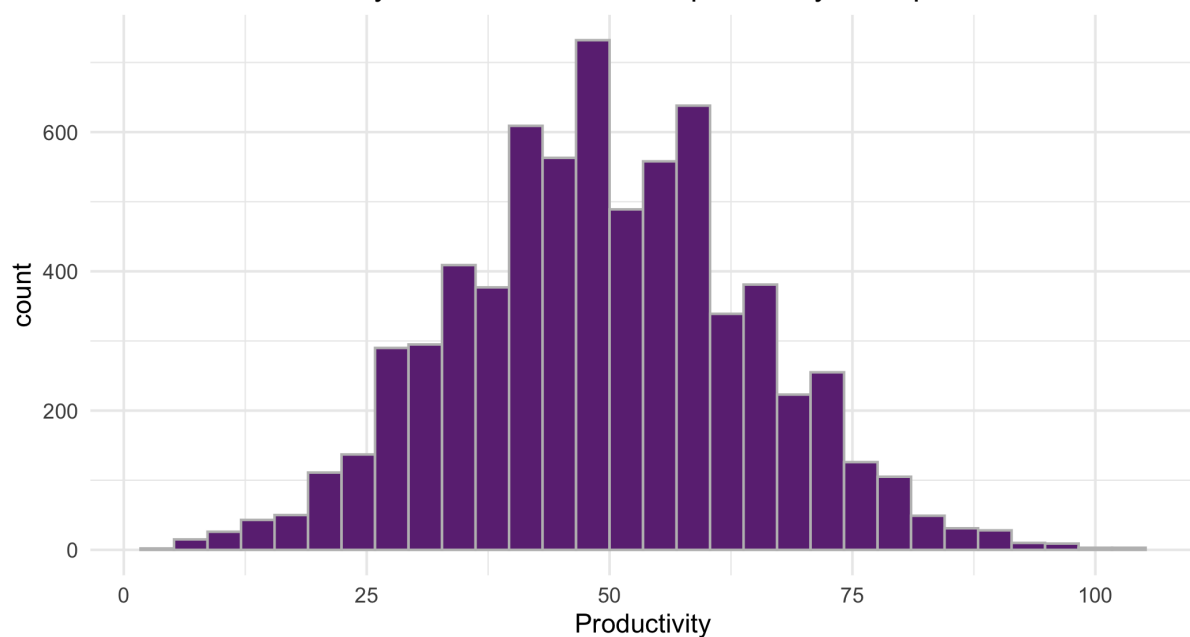
### Research questions

*Use bullet points to 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*

Look! A title. But do you want a title or a caption for your report?



**Informative title for section addressing a research question****Informative title for section addressing a research question****Import Data**

employee\_id, gender, team, financial\_q, role\_seniority, leadership\_for\_level, productivity, salary

applicant\_id (final\_hires)

applicant\_id, team\_applied\_for, cover\_letter, cv, gpa, gender, extracurriculars, work\_experience

applicant\_id, technical\_skills, writing\_skills, speaking\_skills, leadership\_presence

applicant\_id, interviewer\_rating\_1, interviewer\_rating\_2

**Data Cleaning**

Current Employee

Applicants

**Data Visualization**

Part 1 on current employee data findings: 1. fixing quarter, men on avg earn more 2. lower leadership associated with lower productivity and lower salary 3. leadership exceeding expectation does not lead to higher salary 4. productivity follows pyramid shape with role seniorities 5. higher/more senior roles, higher salary 6. for the same productivity, men on avg earn more 7. fixing quarter, across diff levels (entry to manager) men are always the ones receiving highest salary within their team 8. for leadership, almost all “exceed expectation” are men and all “needs improvement” are women 9. for the same role, men earn more 10. for the same team, men earn more

#research question -is there gender bias in current employee remuneration? -fixing team/productivity/quarter/role, men earn more -need improvement: all women, exceed expectation: all men -general info

-lower leadership associated with lower productivity and lower salary -leadership exceeding expectation does not lead to higher salary -productivity follows pyramid shape with role seniorities

-model interest -y:salary x: gender, seniority, team, leadership, productivity, quarter, id -main interest: salary ~ gender -fixed effect: gender, seniority, quarter -random effect: team, leadership, productivity -y: continuous, don't need generalized, -> linear mixed model?

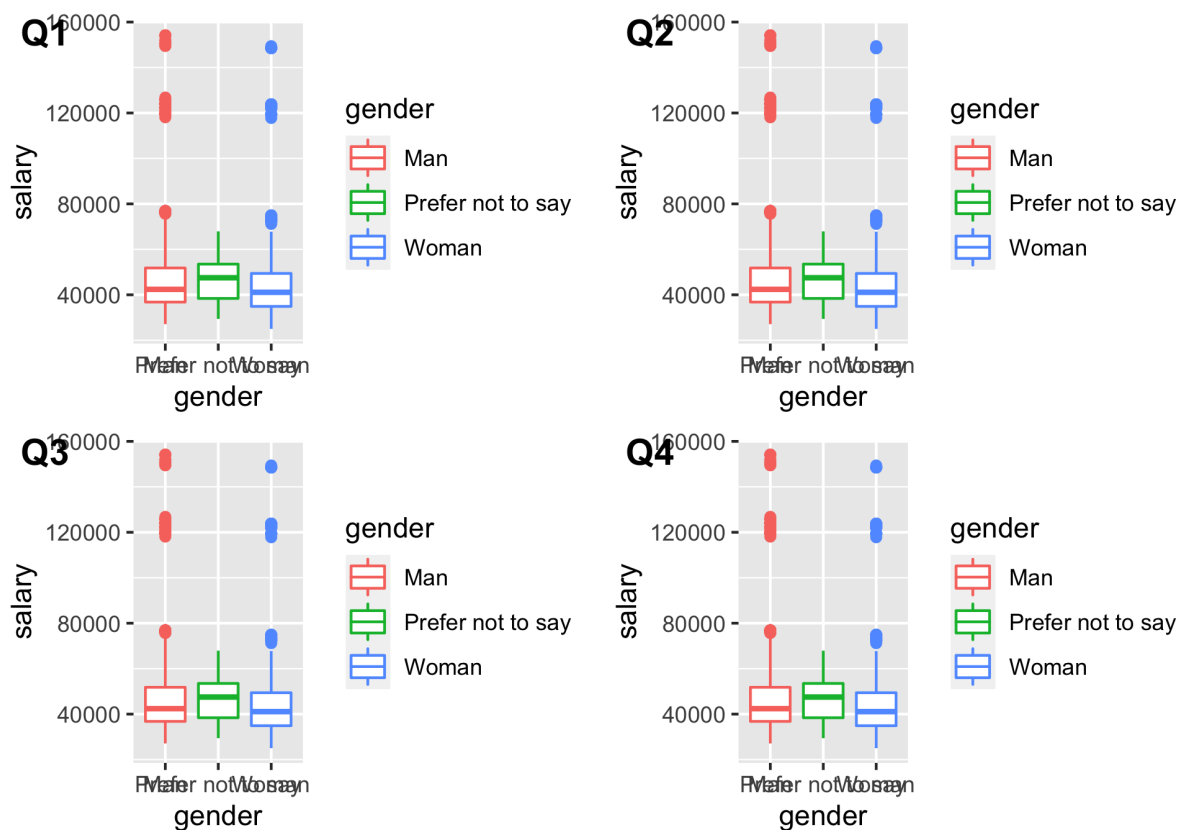
-model breakdown -full model of salary~gender -reduced model (eg. take out productivity)  
-model comparison -use random effect vs. random slope

-other models

-seniority ~ gender

-leadership ~ gender

### Salary distribution for men and women in each quarter



### salary difference across leadership, seniority, and gender for level, fixing productivity

### salary difference in gender across teams, fixing quarter and seniority

Part 2 on hiring data findings: 1.men and women have similar GPA 2.applicants with higher GPA are hired 3.applicants with better skills are hired 4.largest difference in speaking skills, and

least in minimal skills -may suggest bias towards non-native speakers 5.on average, female hires have lower skills than male hires

#research question

-is there gender bias in the hiring process -model: #create phase 2 and phase 3 hired by merging -phase2\_hired ~ gender\* cv gpa cover\_letter (phase1) -phase3\_hired ~ gender\* cv gpa cover\_letter tech writing speaking (phase2) -final\_hired ~ gender\* cv gpa cover\_letter tech writing speaking rating1 rating2 (phase3) -y 0 or 1, not continuous -not linear reg or linear mixed -need generalized linear model -all fixed effect

-reduced model

-final\_hired ~ gender gpa tech writing speaking

-model comparison

-model

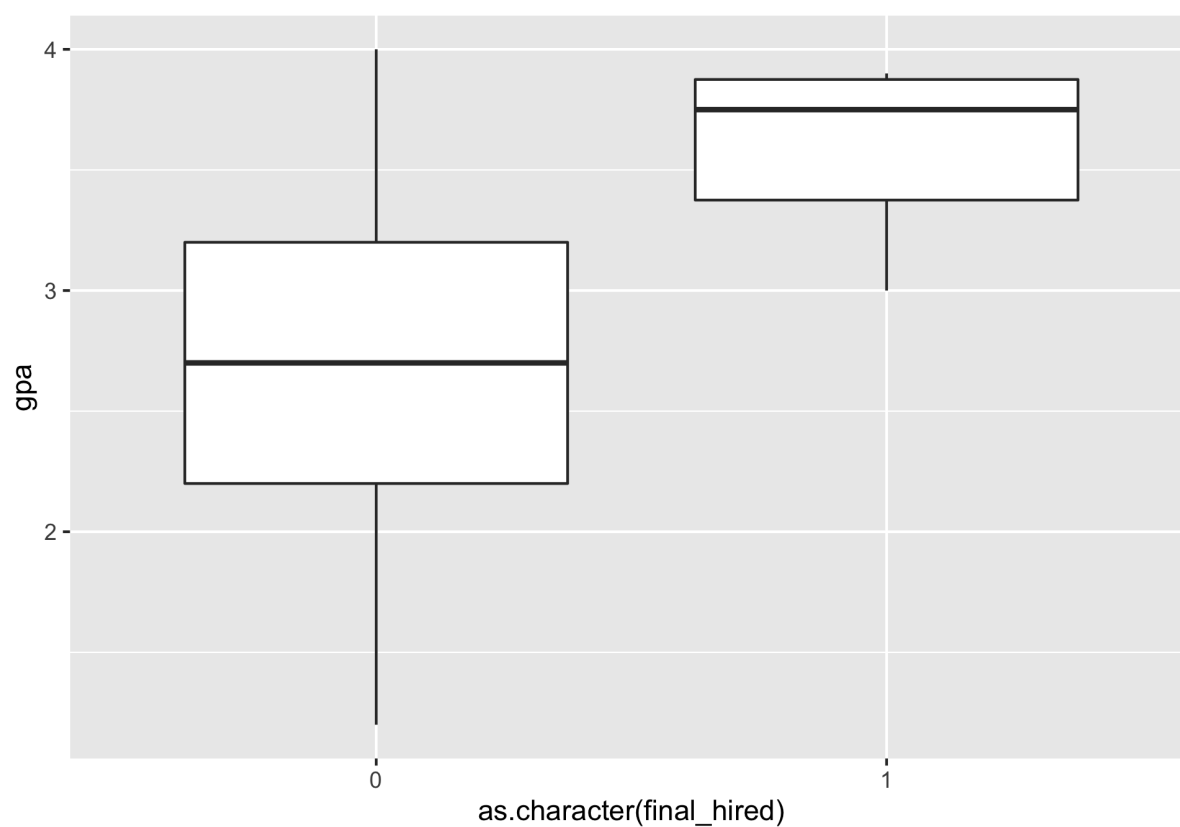
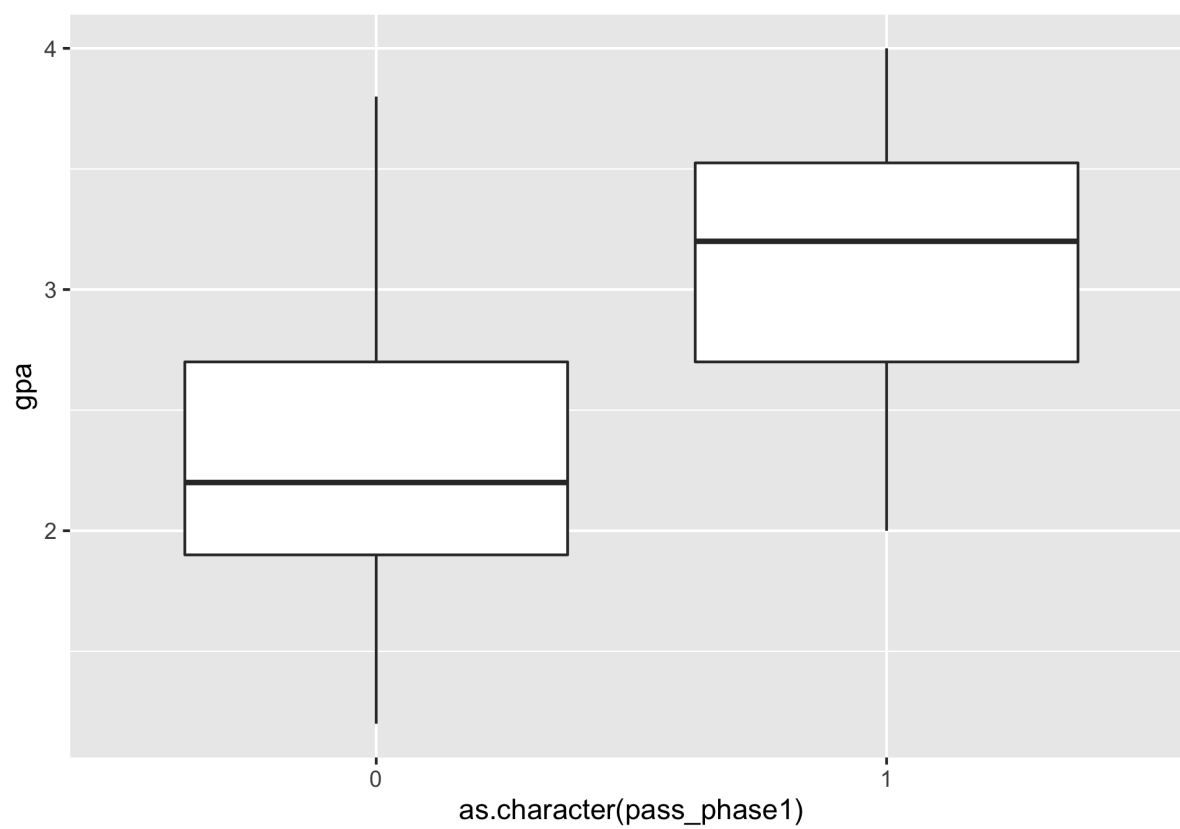
-gpa ~ gender

-skills ~ gender

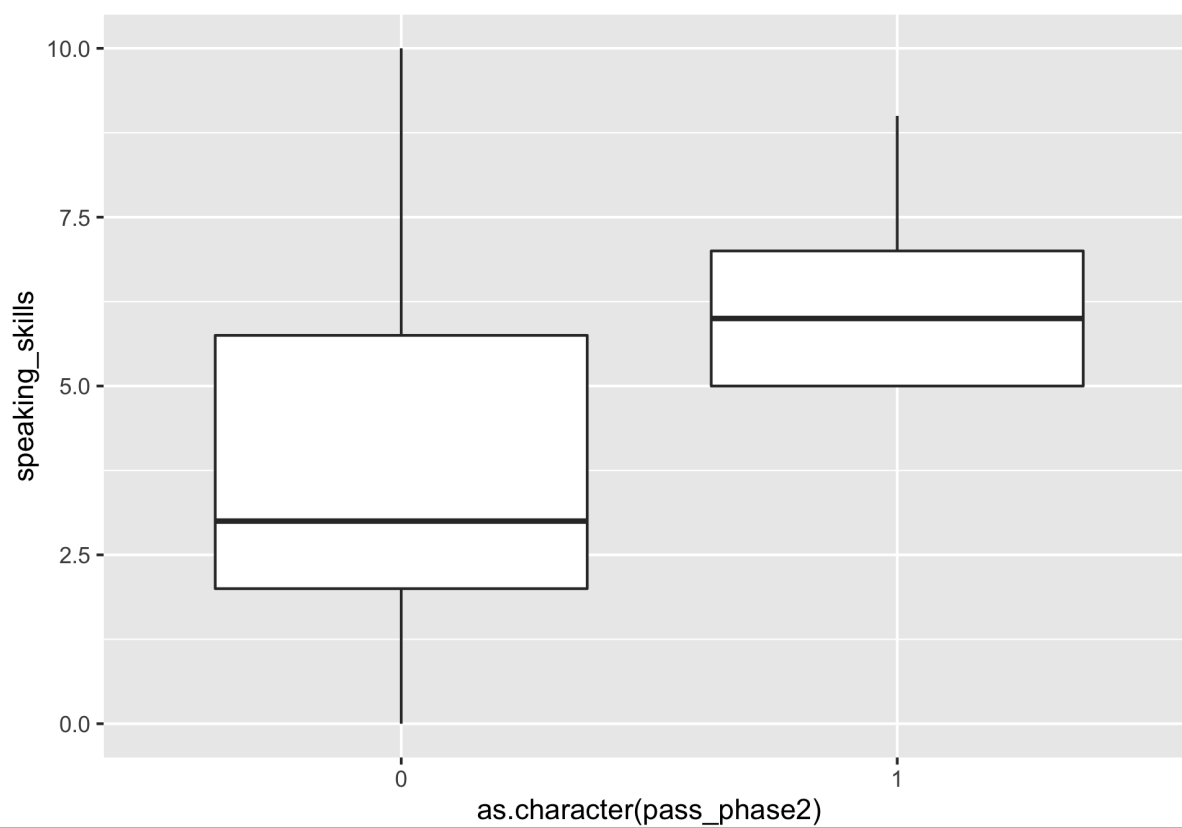
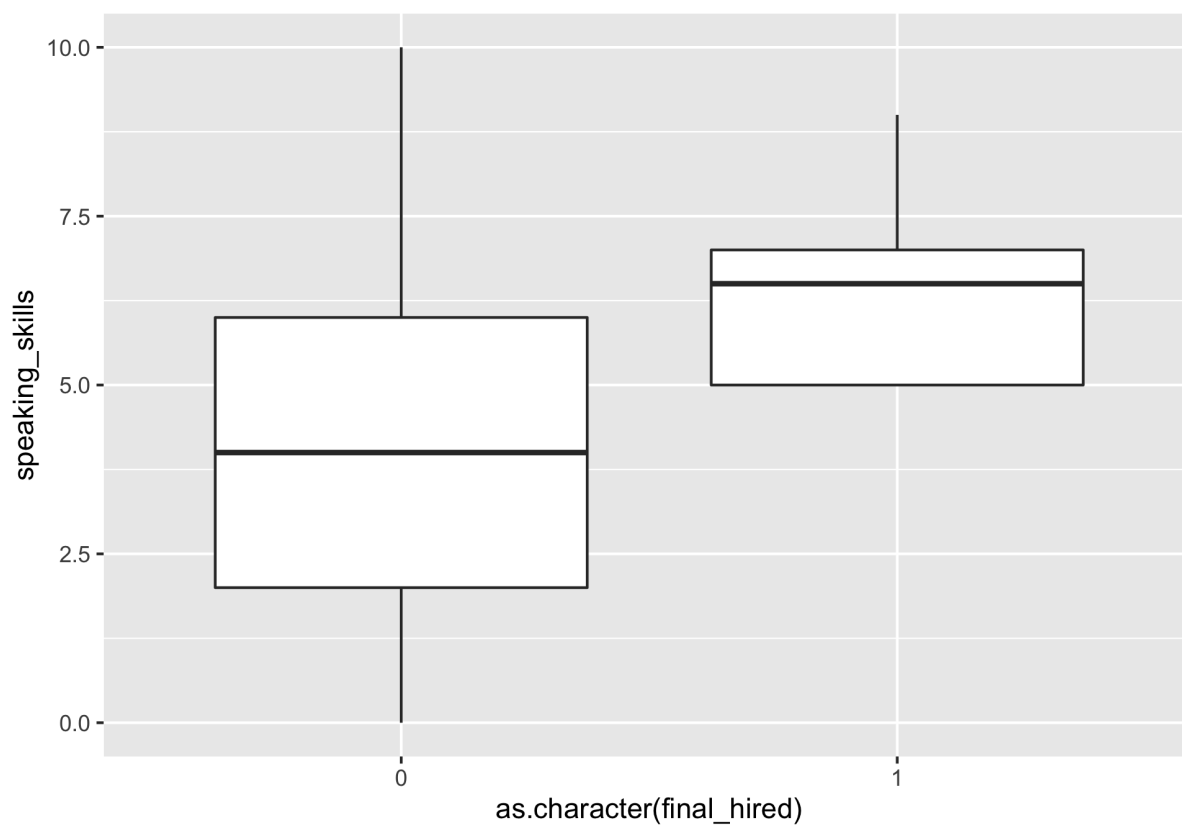
-is there race bias in the hiring process -phase3\_hired ~ speaking\* gender cv gpa cover\_letter tech writing (phase2) -final\_hired ~ speaking\* gender cv gpa cover\_letter tech writing rating1 rating2 (phase3)

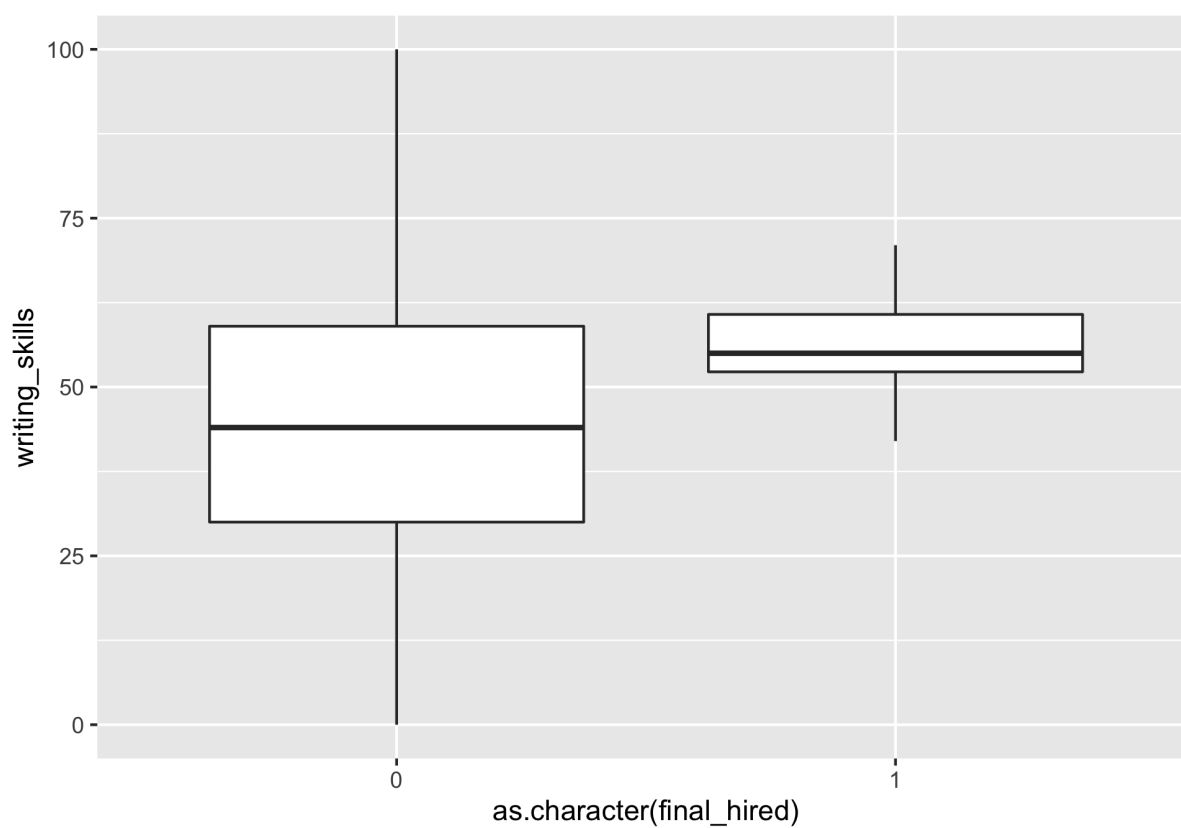


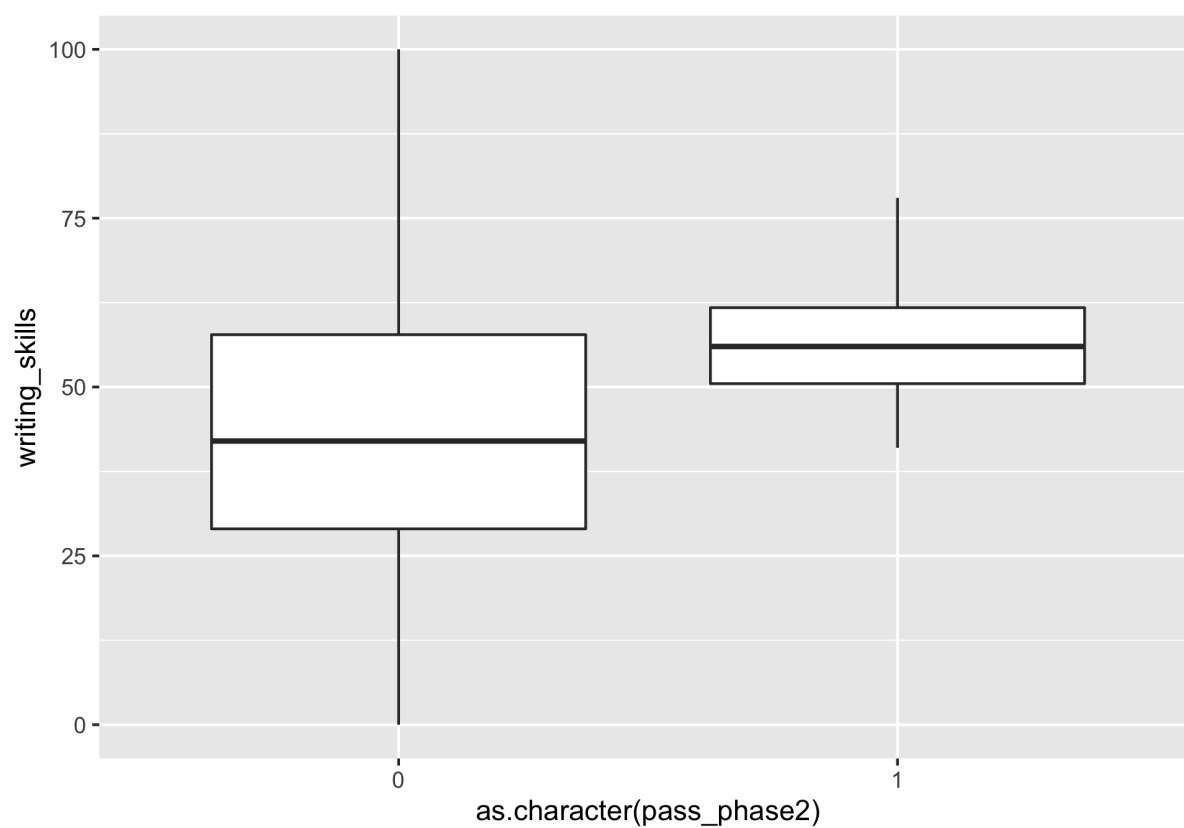


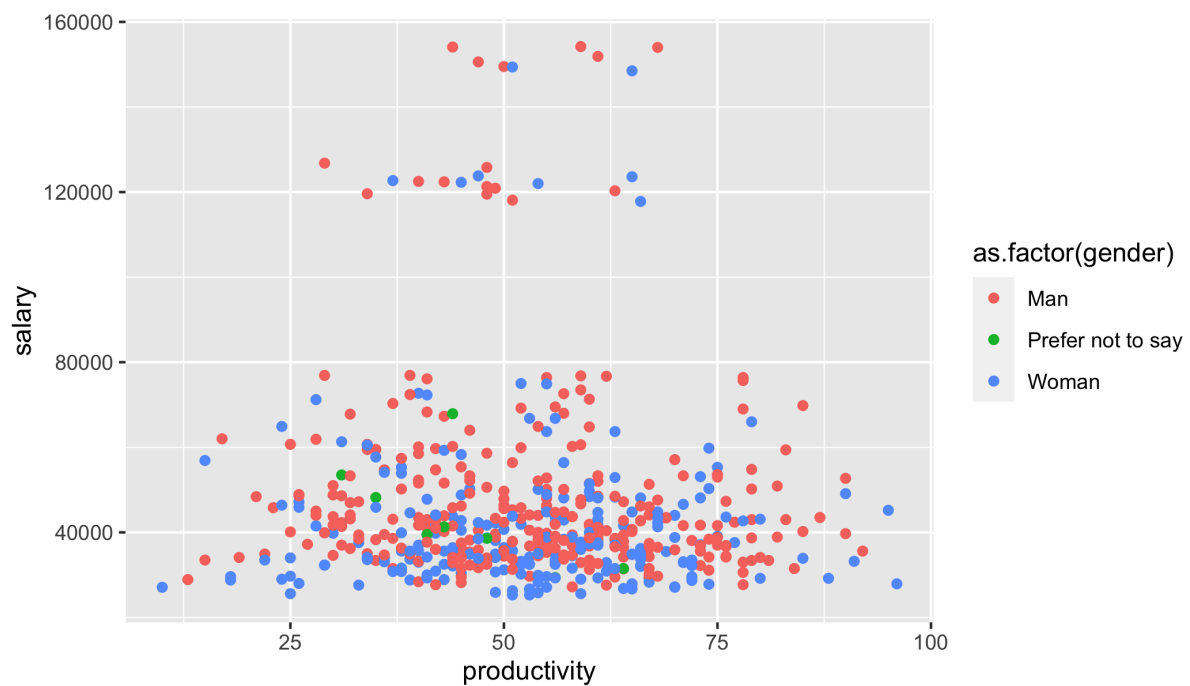
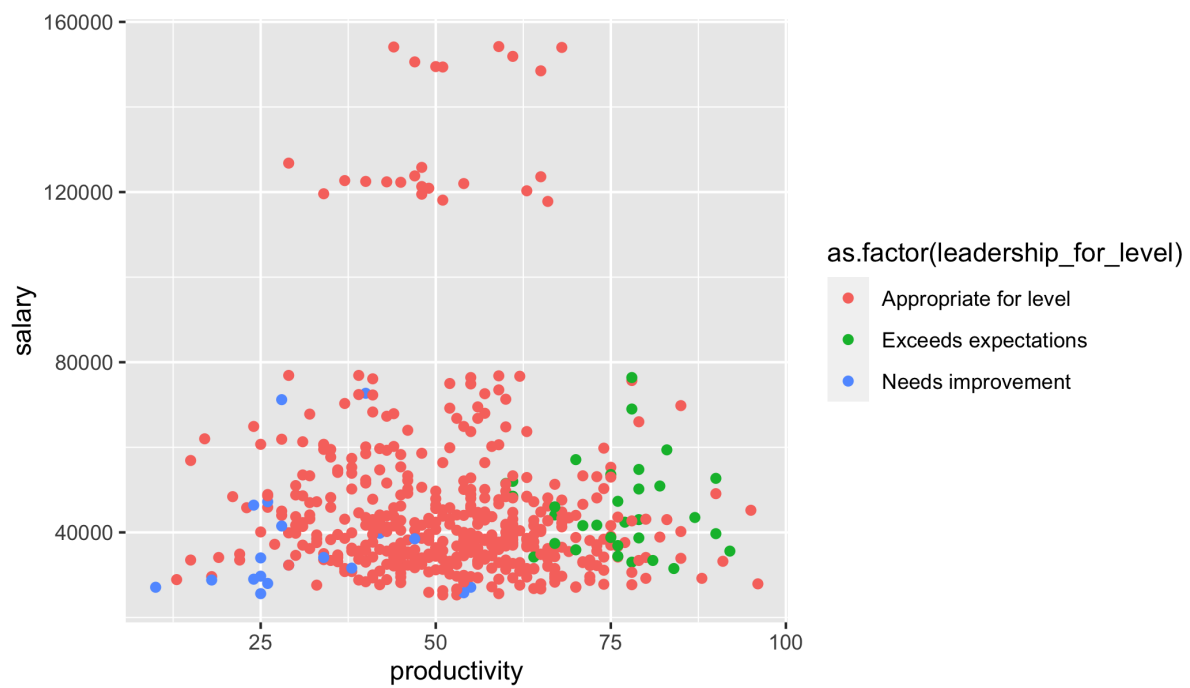
**GPA vs. if hired**



**Speaking skills VS if hired**

**Writing Skills VS if hired**



**Tech Skills VS if hired**

## 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

- Dataset size toooooo small!! especially the final hired data and the phase 3 data (22 observations)

## Consultant information

### Consultant profiles

**Rain Wu.** Rain is a senior consultant with DataOverFlow. She specializes in data visualization. Rain earned her Bachelor of Science, Specialist in Statistics Methods and Practice, from the University of Toronto in 2022. Before joining DataOverFlow, Rain has 3 year of working experience as a data engineer at Aviva in Markham, Toronto.

**Tina Wang.** Tina is a junior consultant with DataOverFlow. She specializes in reproducible analysis. Tina earned her Bachelor of Science, Majoring in Computer Science and Statistics from the University of Toronto in 2022. Tina earned her master degree in financial insurance from the University of Toronto in 2024.

**Yiqu Ding.** Yiqu is a junior consultant with DataOverFlow. She specializes in statistical communication. Yiqu earned her Bachelor of Science, Majoring in Statistics and mathematical application in finance and economics from the University of Toronto in 2022. Yiqu earned her master degree in financial insurance from the University of Toronto in 2024.

### Code of ethical conduct

- We respect and protect confidential data obtained from, or relating to, clients and third parties, as well as personal data and information about employees from Data Over Flow. We only share information when there is a business purpose, and then do so in accordance with applicable laws and professional standards.
- We take proactive measures to safeguard our archives, computers and other data-storage devices containing confidential information or personal data. We promptly report any loss, damage or inappropriate disclosure of confidential information or personal data.
- We use social media and technology in a responsible way and respect everyone we work with. We obtain, develop and protect intellectual capital in an appropriate manner. We respect the restrictions on its use and reproduction.