CS6603 Final Project Report

Olivia Lawson, Sean Nima, Christopher Burgett
olawson6@gatech.edu, snima3@gatech.edu, cburgett3@gatech.edu

1 DATASET SELECTION

1. The dataset selected is the public Home Mortgage Disclosure Act dataset. We have selected the static, snapshot national level level dataset from 2024, and filtered the results down to applicants from the state of Georgia. The link to review the dataset is here[2].

2. The HMDA set belongs to both the credit and housing regulated domains.

3. The HMDA set, filtered to Georgia residents, has 478124 observations, where each observations represents a mortgage loan application. After dropping of zero value data and categories that do not contribute to conclusions (i.e `applicant_sex value 4- Not Applicable`), there are 106195 observations.

4. There are 99 variables within this set. For the purposes of this project, we have isolated 15 variables for analysis.

5. The dependent variables are `interest_rate` and `action_taken`. `action_taken` is a categorical variable with 8 enumerations. We have categorized these enumerations into favorable and unfavorable boolean outcomes, where 1 indicates the loan was received and 0 indicates the loan was denied.`interest_rate` is a continuous, float value. We have categorized this distribution into favorable and unfavorable buckets, determined by considering average the Consumer Financial Protection Bureau's definition of high-cost loans. An interest rate above 7.5% will be considered unfavorable, and below is favorable [1].

6.There are 3 variables in the dataset that are associated with a legally recognized protected class.

- derived_ethnicity: National Origin
- derived_race: Race
- derived_sex : Sex

For this project, we selected race and sex as our protected class variables.

7. All three of the variables mentioned above are protected by the Civil Rights of 1964 act, which prevents discrimination based on race, color, religion, or national

origin. The variable, derived_sex is also protected by the Equal Pay Act of 1963.

2 DATASET EXPLORATION

Our dataset allows multiple applicants for mortgage loan. We limited the number of co-applicants to 2 for this exercise. Due to this, we must accurately demonstrate the permutations of the applicants. This results in 89 possible combinations for race and 9 combinations for sex. Given the number of permutations for race, we have reduced the frequency table to the top 10 most frequent applicant combinations.

1. Identify subgroups

Derived Race (derived_race_new)	Count
White	75441
Black or African American	21188
Asian, Asian Indian	2481
Asian	1224
Asian Indian	829
Asian, Vietnamese	705
Asian, Chinese	599
Asian, Korean	599
Asian, Other Asian	493
Black or African American, White	385

Table 1—Top 10 Raw Frequencies for Derived Race

Sex (derived_sex_new)	**Count**
Male, Female	64782
Female, Male	31856
Female, Female	5153
Male, Male	4063
Applicant selected both male and female, Female	152
Applicant selected both male and female, Male	105
Male, Applicant selected both male and female	42
Female, Applicant selected both male and female	29
Applicant selected both male and female, Applicant selected both male and female	13

Table 2—Top 10 Raw Frequencies for Sex

2. Discretize subgroups

Table 3—Top 10 Derived Race Strings with Encoded Values

Derived Race	Encoded Value
White	75
Black or African American	26
Asian, Asian Indian	15
Asian	8
Asian Indian	9
Asian, Vietnamese	24
Asian, Chinese	17
Asian, Korean	20
Asian, Other Asian	22
Black or African American, White	38

Table 4—Top 10 Derived Sex Strings with Encoded Values

Derived Sex	Encoded Value
Male, Female	7
Female, Male	5
Female, Female	4
Male, Male	8
Applicant selected both, Female	1
Applicant selected both , Male	2
Male, Applicant selected both	6
Female, Applicant selected both	3
Applicant selected both, Applicant selected both	0

3. The selected protected classes are race and sex.

4. Frequency Tables for subgroups versus dependent variables

Table 5—derived_race_new vs favorable_action_taken

Favorable Action Taken	0	1
Race		
Asian	446	778
Asian Indian	353	476
Asian, Asian Indian	816	1665
Asian, Chinese	162	437
Asian, Korean	180	419
Asian, Other Asian	173	320
Asian, Vietnamese	221	484
Black or African American	10085	11103
Black or African American, White	180	205
White	25211	50230

Table 6—derived_race_new vs favorable_interest_rate

Favorable Interest Rate Race	0	1
Asian	613	611
Asian Indian	417	412
Asian, Asian Indian	1041	1440
Asian, Chinese	199	400
Asian, Korean	232	367
Asian, Other Asian	219	274
Asian, Vietnamese	284	421
Black or African American	12537	8651
Black or African American, White	210	175
White	38515	36926

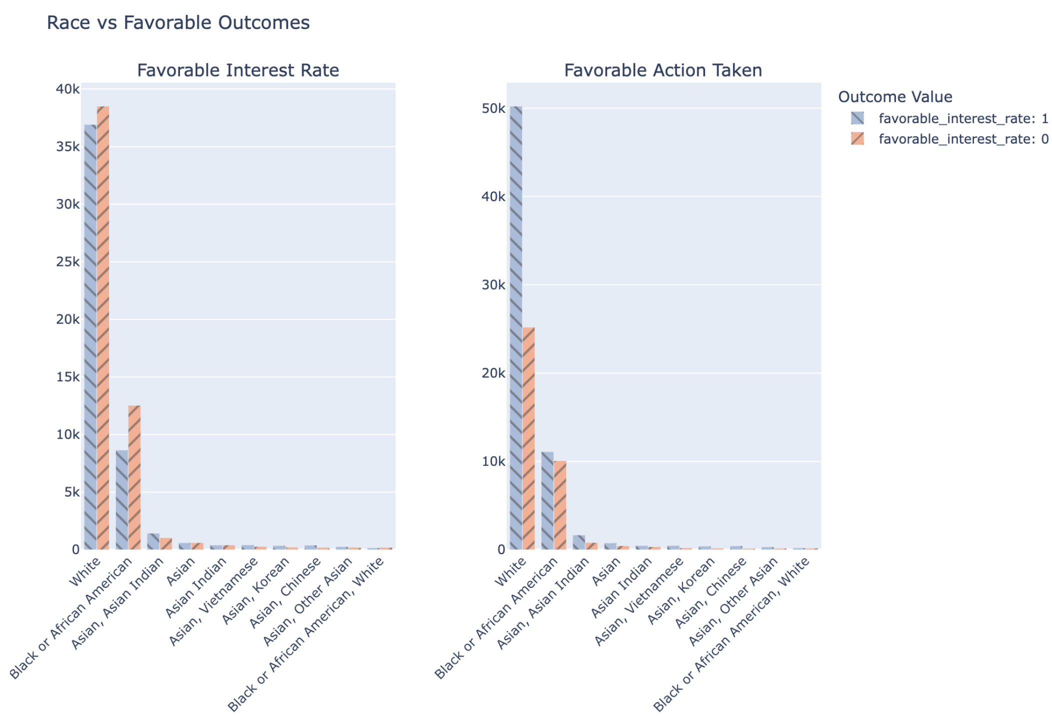
Table 7—derived_sex_new vs favorable_action_taken

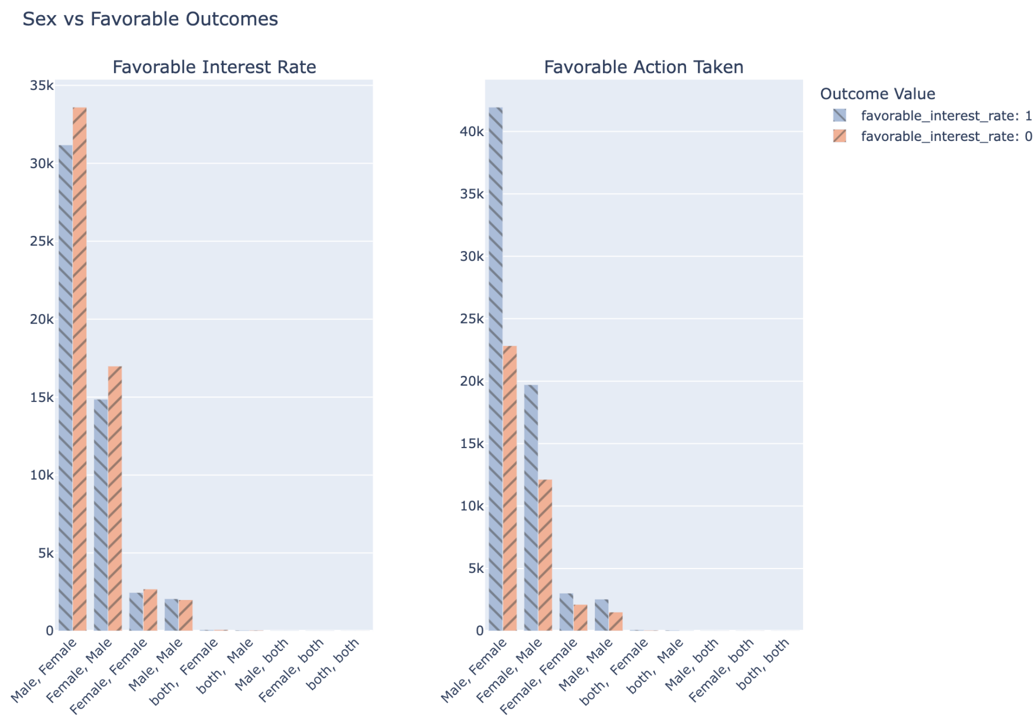
Favorable Action Taken Sex	0	1
Applicant selected both, Applicant selected both	5	8
Applicant selected both, Female	63	89
Applicant selected both, Male	43	62
Female, Applicant selected both	13	16
Female, Female	2125	3028
Female, Male	12134	19722
Male, Applicant selected both	19	23
Male, Female	22844	41938
Male, Male	1513	2550

Table 8—derived_sex_new vs favorable_interest_rate

Favorable Interest Rate Sex	0	1
Applicant selected both, Applicant selected both male and female	7	6
Applicant selected both, Female	79	73
Applicant selected both, Male	51	54
Female, Applicant selected both	15	14
Female, Female	2687	2466
Female, Male	16990	14866
Male, Applicant selected both	23	19
Male, Female	33599	31183
Male, Male	2002	2061

4. Bar Graphs





3 FAIRNESS METRIC AND MITIGATION BIAS

Privileged and Unprivileged group for Sex

- Privileged: Male
- Unprivileged: Female

Privileged and Unprivileged group for Race

- Privileged: White
- Unprivileged: Black/ African American

The two fairness metric algorithms, we selected are **"Statistical Parity Difference"** and **"Disparate Impact"**. The metrics are shown in the table below.

Table 9—Fairness Metrics for Protected Classes and Outcomes (Original Dataset)

Comparison	Statistical Parity Difference	Disparate Impact
Sex vs. Action Taken	0.072747	1.036648
Race vs. Action Taken	0.358704	1.185562
Sex vs. Favorable Interest Rate	-0.011793	0.974679
Race vs. Favorable Interest Rate	-0.084861	0.820610

We decided to apply **reweighting** to our metrics to mitigate bias. The results of the transformed data are in the table below.

Table 10—Fairness Metrics for Protected Classes and Outcomes after Mitigation Strategies

Comparison	Statistical Parity Difference	Disparate Impact
Sex vs. Action Taken	0.042277	1.021186
Race vs. Action Taken	0.195128	1.099385
Sex vs. Favorable Interest Rate	-0.003563	0.992303
Race vs. Favorable Interest Rate	-0.039468	0.915164

4 MITIGATING BIAS

4.1 Original Dataset

Table 11—Fairness Metrics on Model Predictions

Comparison	Statistical Parity Difference	Disparate Impact
Sex vs. Prediction	1.386294	1.352153
Race vs. Prediction	0.891242	1.206820

4.2 Transformed Dataset

Table 12—Fairness Metrics on Model Predictions

Comparison	Statistical Parity Difference	Disparate Impact
Sex vs. Prediction	0.007303	1.007289
Race vs. Prediction	0.002078	1.002078

4.3 Comparison Table

Table 13—Summary of Fairness Metric Changes for Sex

(a) Disparate Impact

Stage	Disparate Impact	Change Compared to Previous
Original Dataset	1.036648	NA
After Transforming Dataset	1.021186	Positive
After Training Classifier on Original	1.352153	Negative
After Training Classifier on Transformed	1.007289	Positive

(b) Statistical Parity Difference

Stage	Statistical Parity Difference	Change Compared to Previous
Original Dataset	0.072747	NA
After Transforming Dataset	0.422770	Negative
After Training Classifier on Original	1.386294	Negative
After Training Classifier on Transformed	0.007303	Positive

5 ANALYSIS

5.1 Team Members

Olivia Lawson, Chris Burgett, Sean Nima

Table 14—Summary of Fairness Metric Changes for Race

(a) Disparate Impact

Stage	Disparate Impact	Change Compared to Previous
Original Dataset	1.185562	NA
After Transforming Dataset	1.099385	Positive
After Training Classifier on Original	1.206820	Negative
After Training Classifier on Transformed	1.002078	Positive

(b) Statistical Parity Difference

Stage	Statistical Parity Difference	Change Compared to Previous
Original Dataset	0.358704	NA
After Transforming Dataset	0.195128	Positive
After Training Classifier on Original	0.891242	Negative
After Training Classifier on Transformed	0.002078	Positive

5.2 Graphs

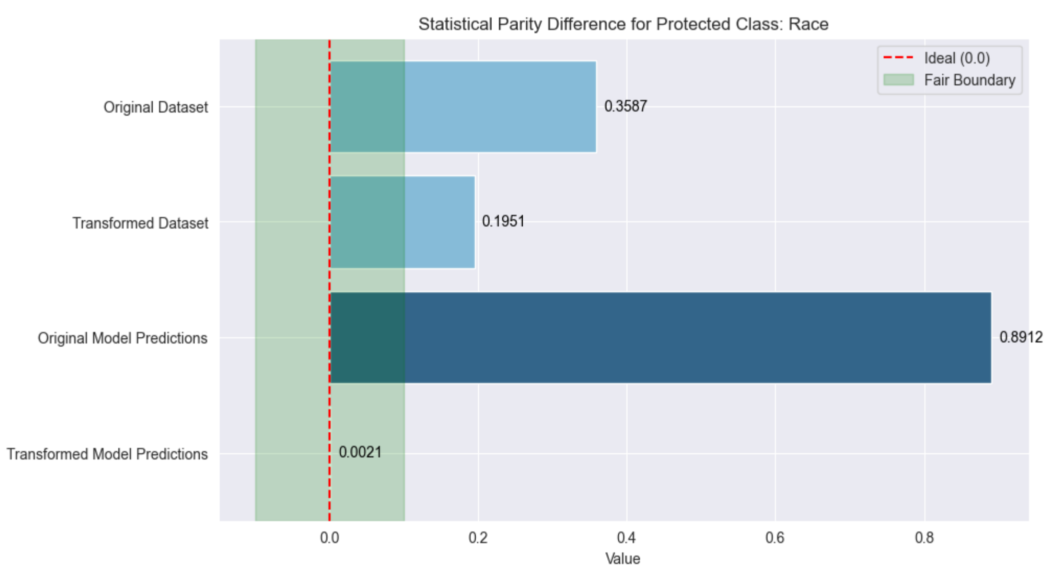


Figure 1—Comparison of Statistical Parity Difference for Race Before and After Mitigation.

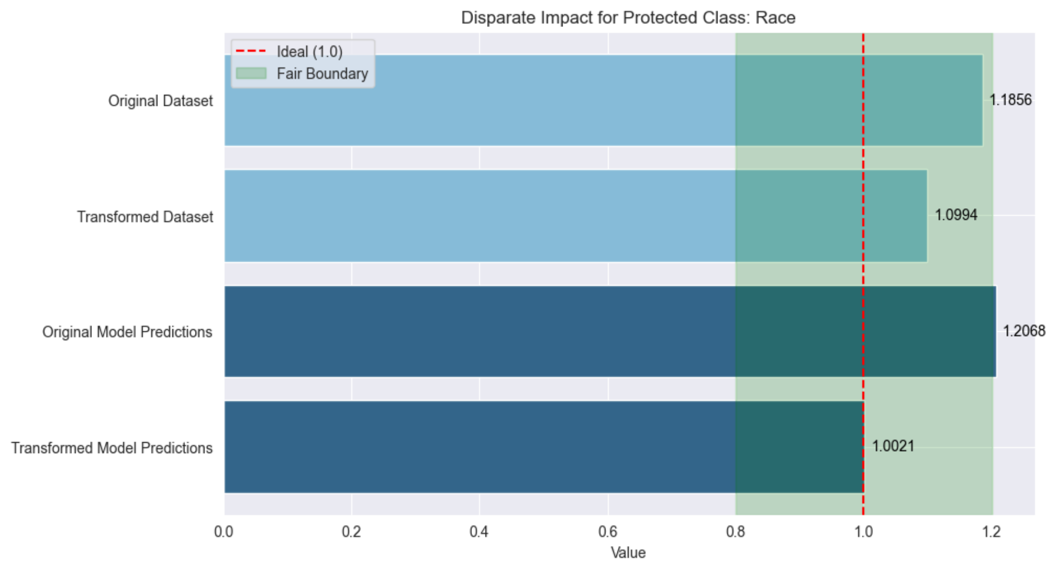


Figure 2—Comparison of Disparate Impact for Race Before and After Mitigation.

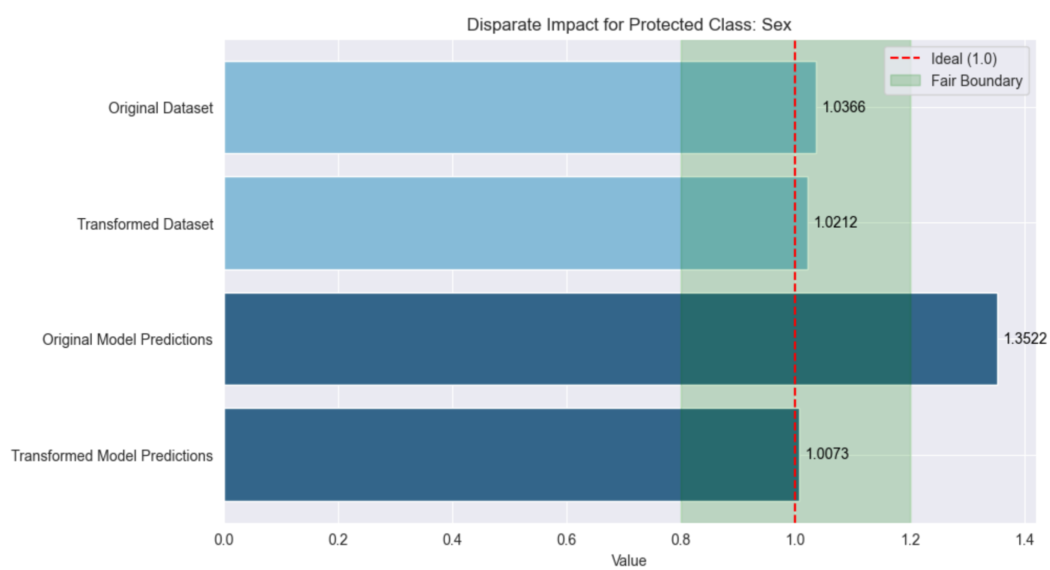


Figure 3—Comparison of Disparate Impact for Sex Before and After Mitigation.

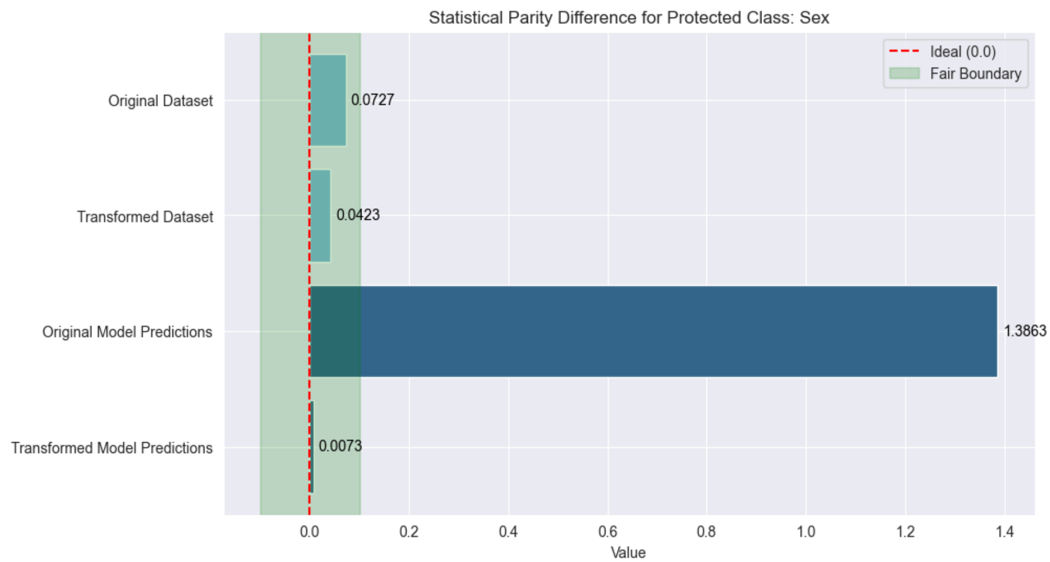


Figure 4—Comparison of Statistical Parity Difference for Sex Before and After Mitigation.

5.3 Best Fairness Metric

While both fairness metrics were effective in this project, Disparate Impact is the superior choice. Given that the analysis is in the highly regulated Credit and Housing domain, Disparate Impact is the most appropriate metric because it directly aligns with the legal and regulatory standards used to ensure that the rate of favorable outcomes is not substantially lower for any protected group.

5.4 Olivia's Analysis

The mitigation methods reduced bias across both fairness metrics. Sex statistical parity difference fell from 1.386 to 0.0073, and disparate impact moved from 1.352 to 1.0073, indicating a shift toward parity. Race statistical parity improved from 0.891 to 0.0021, and disparate impact rose from 0.7789 to 1.0021, correcting the initial disadvantage faced by Black or African American applicants. After transformation, neither group showed a clear advantage, and fairness metrics stayed near ideal values. There are risks with using this bias mitigation approach, however. Though we analyzed one disadvantaged group with respect to race, there are other groups with less representation in the dataset who may negatively affected by the mitigation technique. For example, Asian groups may be hurt in this process as they were not optimized for.

5.5 Chris's Analysis

The reweighting approach was highly effective at mitigating bias for the selected groups. Initially, the model clearly favored the privileged groups. For example, the disparate impact score of 1.3522 for sex shows that males received favorable outcomes at a much higher rate. The reweighting method corrected this imbalance, bringing the final model's DI for sex to 1.0073 and its statistical parity difference to 0.0073, both close to their ideal fairness targets. This approach successfully removed the disadvantage faced by the unprivileged groups (Female and Black/African American) without disadvantaging the privileged groups in the final outcome. The primary issue with this method is its narrow focus. With this process being optimized specifically for a binary comparison between White and Black/African American applicants, it risks negatively affecting other underrepresented groups, such as Asian applicants, who were not accounted for in the reweighting. This raises a significant ethical concern where we achieved fairness for our target groups, but may have inadvertently ignored or worsened biases against other minority populations.

5.6 Sean's Analysis

The reweighting approach we implemented proved highly effective at mitigating bias for both protected classes. After applying this technique and training a new model, we observed improvement in fairness metrics. For race, the Statistical Parity Difference (SPD) and Disparate Impact (DI) reached near-ideal values of 0.002078 and 1.002078, respectively. Similarly, for sex, the model achieved an SPD of 0.007303 and a DI of 1.007289. Regarding which groups received an advantage, the initial model trained on the original data appeared to favor the unprivileged groups. After mitigation, the final model did not disadvantage any of the analyzed groups, as evidenced by the nearly perfect fairness scores. A key issue with this method is that it acts as a technical patch rather than addressing the root cause of the bias. While our model is now fair, it does not solve the underlying systemic issues that created the initial biases in the data.

6 REFERENCES

[1] Bureau, Consumer Financial Protection (2024). *12 CFR § 1026.32 - Requirements for high-cost mortgages*. https://www.consumerfinance.gov/rules-policy/regulations/1026/32/. Accessed: 2025-07-17.

[2] Council, Federal Financial Institutions Examination (2024). *Snapshot National Loan-Level Dataset*. `https://ffiec.cfpb.gov/data-publication/snapshot-national-loan-level-dataset/2024`. Accessed: 2025-07-17.

7 APPENDICES

You may optionally move certain information to appendices at the end of your paper, after the reference list. If you have multiple appendices, you should create a section with a *Heading 1* of "Appendices." Each appendix should begin with a descriptive *Heading 2*; appendices can thus be referenced in the body text using their heading number and description, e.g. "Appendix 5.1: Survey responses." If you have only one appendix, you can label it with the word "Appendix" followed by a descriptive title, e.g., "Appendix: Survey responses."

These appendices do not count against the page limit, but they should not contain any information required to answer the question in full. The body text should be sufficient to answer the question, and the appendices should be included only for you to reference or to give additional context. If you decide to move content to an appendix, be sure to summarize the content and note it in relevant place in the body text, e.g., "The raw data can be viewed in *Appendix 5.1: Survey responses*."