



CITY OF SUGAR LAND

# HOW COMPUTER SCIENCE CAN ENHANCE DATA

Ralph B. Louis, Erik Schenck



## Introduction

For any business or organization, one of the most important factors to its success is its data collection. Data is like the immune system of an organization because it is used to indicate if there are any inefficiencies and in which direction, they are heading in. Gathering the data, however, is only the first, and usually the easiest, step. In an article published by the “Harvard Business Review”, they discuss how the more data a company has, the higher risk that it has to be unhelpful or makes the path forward more unclear. It is because data, in and of itself, is useless. In order for it to be used to inform decisions, it must be “distilled” to a more consumable form.

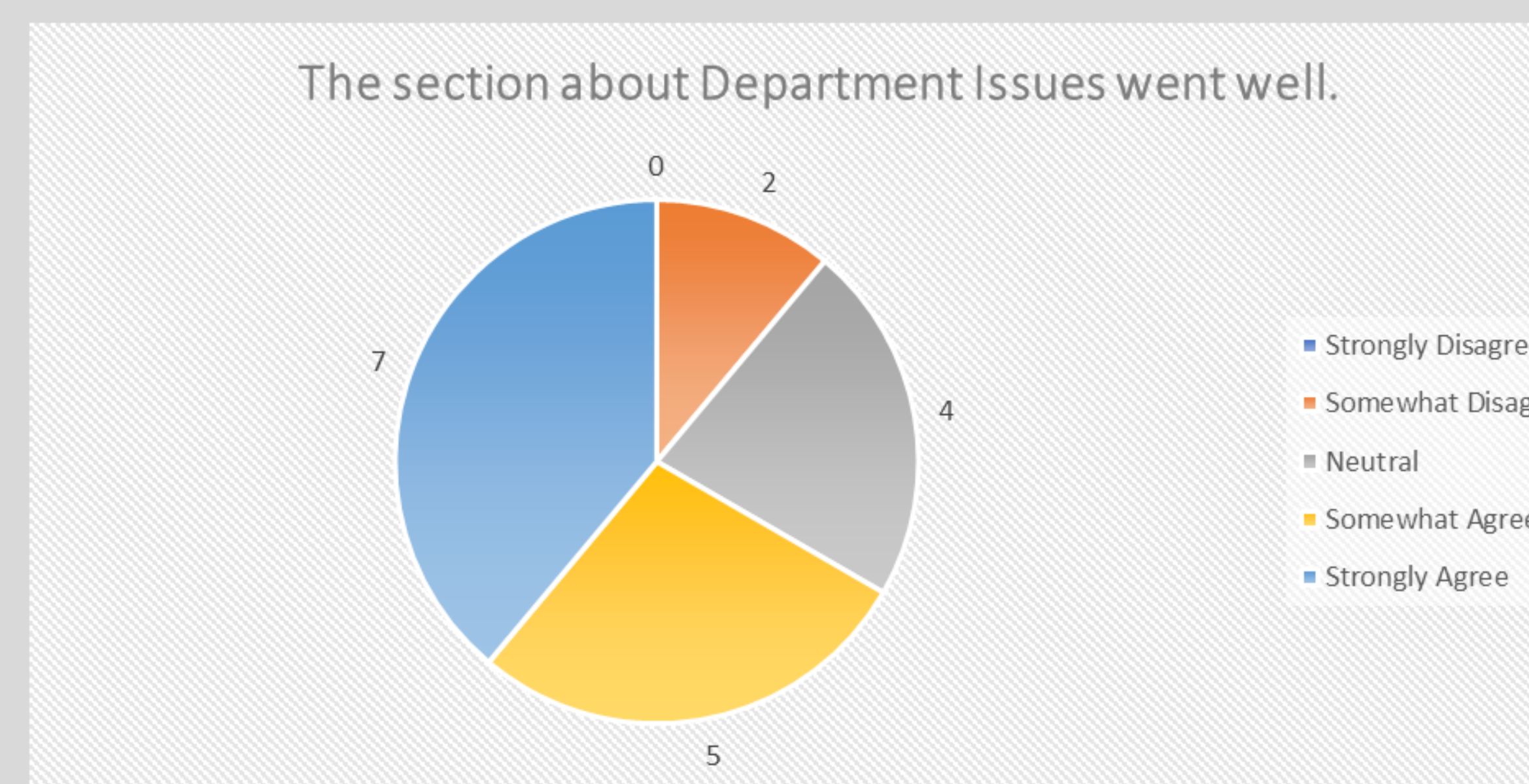
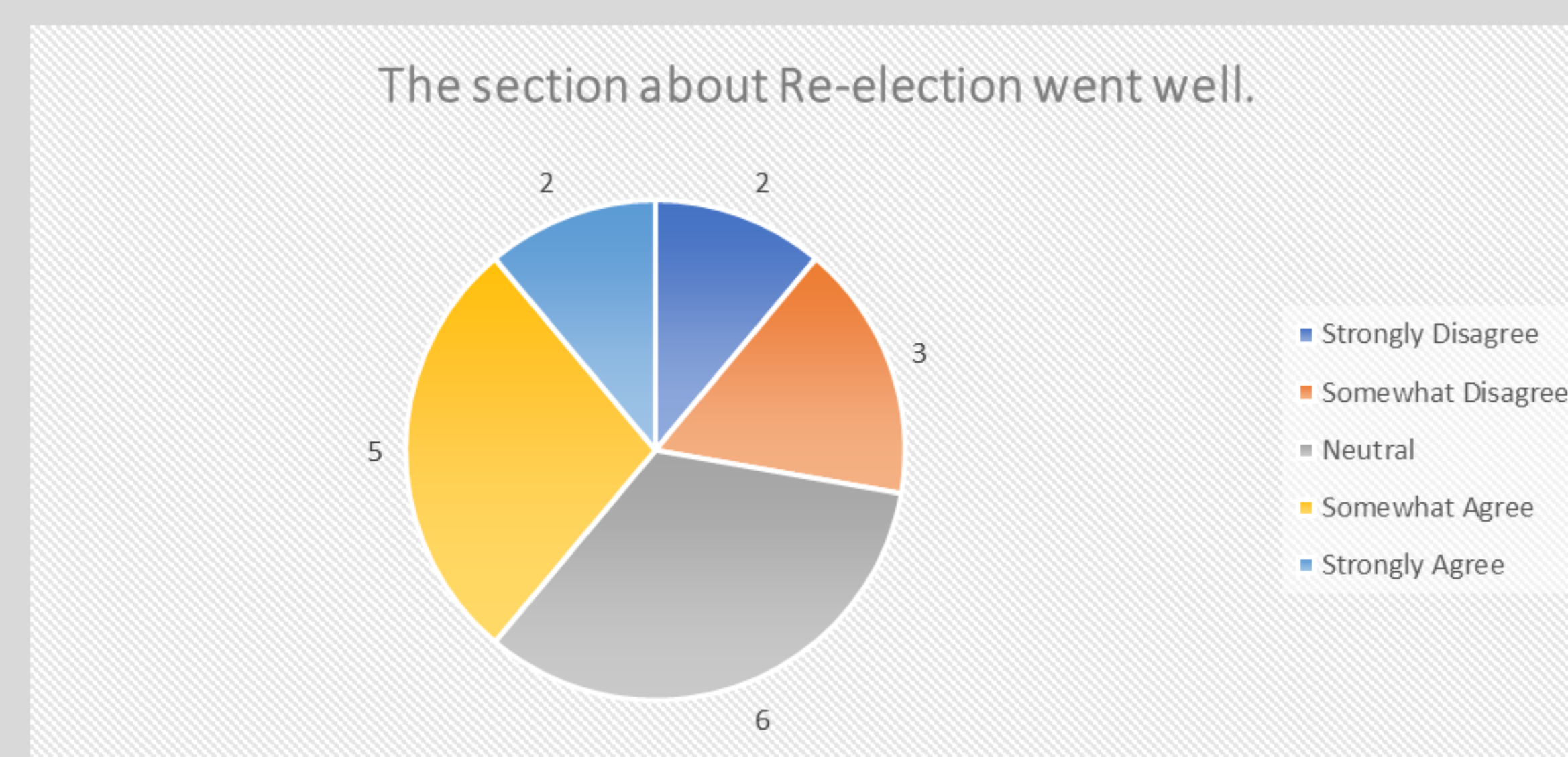
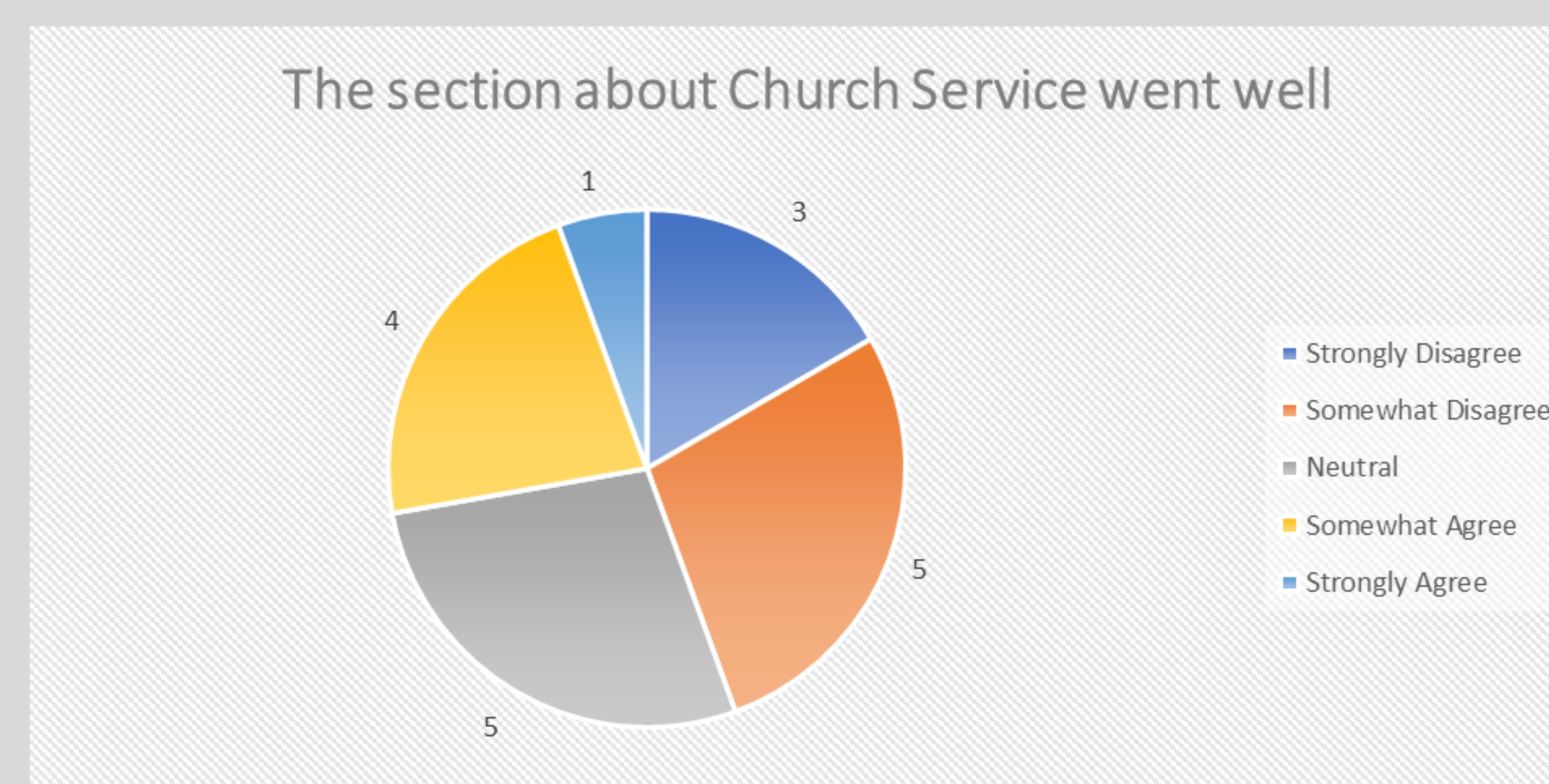
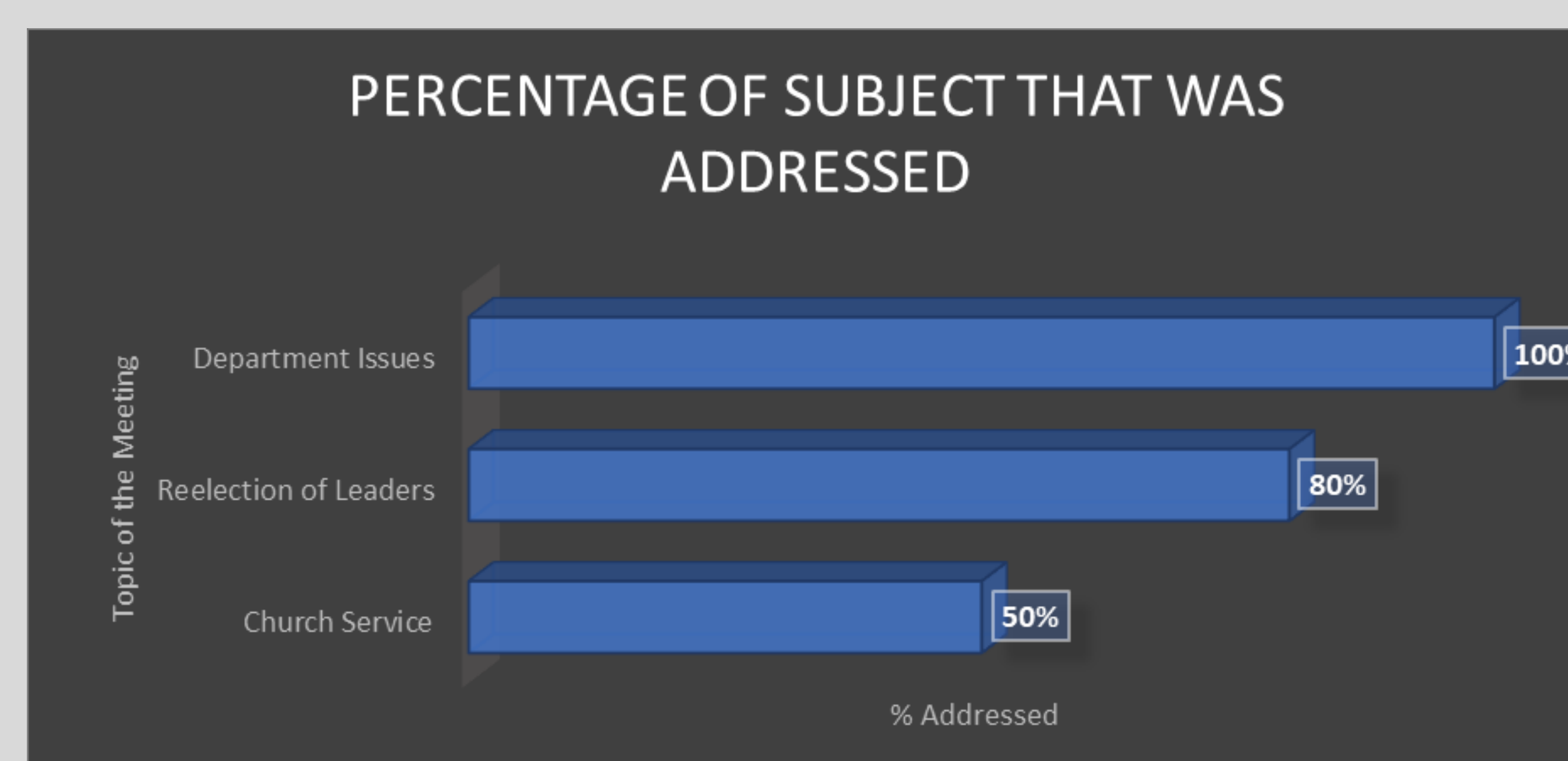
Moreover, in a Forbes article, they recognize that data shouldn’t just be simplified, but put into an “easily understandable form suited for decision-making”, The next logical question is how best is this “form” achieved? Therefore, this experiment is meant to see whether using a created computer program to clean up data is more effective for action than pre-built methods like Excel.

## Methodology

In order to test my hypothesis, the investigation process will be split into two parts. The first step will be to conduct a committee meeting with the leaders of Houston Haitian Bethany SDA church. During the meeting, there will be three types of data presentation for them to discuss: “raw” data, data cleaned by Excel, and data simplified with a computer program (In that order). At the end of each section's time, record the perceived percentage of the topic covered in the timeframe for that section.

Secondly, there will be a conclusion survey handed out to understand their opinions on the flow of the entire meeting and how the data helped advance or hinder decisions. This is used to see the meeting from their point of view, instead of from the person conducting the experiment.

## Results



## Conclusion

The data from both the surveys and my own observations seem to suggest that the data that was cleaned up by a computer program worked better in leading to meaningful discussion and actionable decisions to take place. In Figure 1, there is a trend that the longer the meeting went on, the more the subject was addressed in fullness. Now, to clarify, the data collected for Figure 1 only considered the points made within the timeframe set and not the entirety of the section. Figures 2 – 4 also reflect this summary, as the “Strongly Disagree” and “Somewhat Disagree” decreased from the first section to the last section. The “Somewhat Agree” and “Strongly Agree” increased from the first section to the last section as well.

Not included in the Results section were the few comments left by the participants, which fleshed out their thoughts a little bit more. These comments showed that they believed that the first section felt “too slow” and “very repetitive” while the last section was “very quick and smooth”.

## Discussion

One thing that must be considered when comparing the pre-built graphs of Excel to my computer-programmed design is that going the computer science route is much harder to maintain. That is because each data set is usually requiring its own code for it to be 100% effective to that data. In the survey and written responses, it was seen as most neutral, while having the benefit of being a “One size fits all” type of application. Moreover, there could a little bias in the responses as they knew that I was using the meeting for my experiment. It could be the case that without them knowing, the difference between the sections of the meeting, especially between the 2nd and 3rd ones, would be massively under-noticed.