

# Problem Set 2 - pdf Submission

Submit Assignment

**Due**

Monday by 11:59pm

**Points**

40

**Submitting**

a file upload

**File Types**

pdf

**Available**

Oct 15 at 12am - Oct 20 at 1:59am

5 days

Please submit your final Report here. The report should be in a .pdf format.

This document will be formally graded.

Problem Set 2 - Rubric

Criteria	Ratings				
Title - An informative title is included in the PDF. - Tell the reader what your story is - don't waste their time. - Ideally tell them what happens at the end of the story. - "Problem Set X" is NOT an informative title.	2.0 pts Full Marks Exceptional		1.0 pts There is a title		0.0 pts No Title
Authors, Data, Repo The author and date are clearly included under the title. A link to a GitHub repo is also contained (likely, but not necessarily, through a statement in the report such as: 'Code and data supporting this analysis is available at: X').	2.0 pts Full Marks		0.0 pts No Marks		
Abstract - An abstract is included and appropriately pitched to a general audience. - The abstract answers: what was done, what was found, and why this matters (all at a high level). - If your abstract is longer than four sentences then you need to think a lot about whether it is too long. It may be fine (there are always exceptions) but you should probably have a good reason.	4.0 pts Exceptional Full and complete summary. Exceptional flow and communication.	3.0 pts Great Great - an aspect is poorly done or tone is slightly inappropriate		2.0 pts Adequate Adequate, but there are still major issues.	1.0 pts Poor Does not achieve the basic an executive summary.
Introduction - The introduction is self-contained and tells a reader everything they need to know, including putting it into a broader context. - Your introduction should provide a bit of broader context to motivate the reader, as well as providing a bit more detail about what you're interested in, what you did, what you found, why it's important, etc. - A reader should be able to read only your introduction and have a good idea about the research that you carried out. - It would be rare that you would have tables or figures in your introduction (again there are always exceptions but think deeply about whether yours is one). - It must outline the structure of the report.	4.0 pts Exceptional Introduction is complete and is well written.	3.0 pts Great Overall complete, but there are still aspects that could use improvement.	2.0 pts Adequate The introduction is complete, but there is still major room for improvement (either in terms of flow, communication and/or content).		1.0 pts Poor The introduction is included but is severely lacking. The introduction is missing key components and there is room for improvement in overall communication and

Criteria	Ratings				
<p><b>Data</b></p> <p>When you discuss the dataset (in the data section) you should make sure to discuss at least:</p> <ul style="list-style-type: none"> <li>- The source of the data.</li> <li>- The methodology and approach that is used to collect and process the data.</li> <li>- The population, the frame, and the sample.</li> <li>- Information about how respondents were found. What happened to non-response?</li> <li>- What are its key features□ strengths□ and weaknesses about the survey generally.</li> <li>- This is just some of the issues strong submissions will consider. Show off your knowledge. If this becomes too detailed then you should push some of this to footnotes or an appendix.</li> <li>- You should thoroughly discuss the variables in the dataset that you use. Are there any that are very similar that you nonetheless don't use? Did you construct any variables by combining various ones?</li> <li>- What do the data look like?</li> <li>- Plot. The. Raw. Data. (or as close as you can get to it).</li> <li>- Exceptional means that when I read your submission I learn something about the dataset that I don't learn from any other submission (within a reasonable measure of course).</li> </ul>	<p><b>8.0 pts</b> <b>Exceptional</b></p>	<p><b>7.0 pts</b> <b>Great</b></p> <p>Missing some minor components and is well written. Or is complete, but could use some minor editing.</p>	<p><b>6.0 pts</b> <b>Good</b></p> <p>Mostly fine in general, but includes more than one minor issue.</p>	<p><b>4.0 pts</b> <b>Adequate</b></p> <p>Gets the job done, but a lot of shortcomings.</p>	<p><b>2.0 pts</b> <b>Major Issues</b></p> <p>Missing lots of compo and needs major improvements.</p>

Criteria	Ratings							
<b>Model</b> - The model is well-explained, justified, and appropriate. - When you discuss your model you must be extremely careful to spell out the statistical model that you are using defining and explaining each aspect and why it is important. - The model needs to be written out in appropriate mathematical notation but also in plain English. - Every aspect of that notation must be defined otherwise the most this section can receive is poor. - If the model is Bayesian, then priors need to be defined and sensible. - Discussion needs to occur around how features enter the model and why. For instance, (and these are just examples) why use ages rather than age-groups, why does province have a levels effect, why is gender categorical, etc? - In general, in order to be adequate, there needs to be a clear justification that this is the model for the situation. - Alternative models, or variants, must be discussed and strengths and weaknesses made clear. - You should mention the software that you used to run the model. - You should be clear about model convergence, model checks, and diagnostic issues. - Great answers would discuss things such as, how do the sampling and survey aspects that you discussed assert themselves in the modelling decisions that you make. Again if it becomes too detailed then push some of the details to footnotes or an appendix.	<b>10.0 pts Exceptional</b> Complete and exceptionally communicated.	<b>8.0 pts Great</b> This section only contains one or two minor issues, but overall well done. Good communication and well thought out.	<b>6.0 pts Good</b> Contains one major issue, or a lot of minor issues. Overall, meets most of the criteria in the section. Could use improvement in communication and flow.	<b>5.0 pts Fine</b> Contains one major issue and some minor issues. Section is completed, but does not meet all criteria. Communication could use improvement.	<b>4.0 pts Adequate</b> Contains multiple major issues and/or many minor issues. Section is completed, but does not meet all criteria. Communication could use improvement.	<b>3.0 pts Barely Adequate</b> Contains many major and minor issues. Section is does not meet all criteria. Communication could use improvement.	<b>2.0 pts Not Great</b> Contains many major and minor issues. Section is does not meet all criteria. Communication could use major improvement.	<b>1.0 pts Poor</b> Contains major and minor issues. Section is clearly incomplete. Communication does not meet all criteria. Communication is poorly communicated.
<b>Results</b> - Results will likely require summary statistics, tables, graphs, images, and possibly statistical analysis or maps. - Show the reader the results by plotting them. Talk about them. Explain them. That said, this section should strictly relay results.	<b>6.0 pts Exceptional</b> Complete and well communicated.	<b>5.0 pts Great</b> Complete and well communicated, with one or two minor issue(s).	<b>4.0 pts Good</b> Complete with one major issue or multiple minor issues.	<b>3.0 pts Good</b> Complete with at most 2 major issues or a lot of minor issues.	<b>2.0 pts Adequate</b> Completed, but missing components. Lots of major issues.	<b>1.0 pts Poor</b> Incomplete and missing components. Lots of major issues.		

Discussion Criteria	8.0 pts Exceptional	7.0 pts Great	6.0 pts Good	5.0 pts Fine	4.0 pts Adequate	2.0 pts Almost Adequate	1.0 pts Poor	0.0 No
likely require summary statistics, tables, graphs, images, and possibly statistical analysis or maps. - Show the reader the dataset by plotting the data. Talk about it. Explain it. It is this dataset that will allow you to tell your story, so if the reader isn't comfortable with it then they will be difficult to convince. - Explain where the dataset came from, the bias that may exist, and what you did to adjust for that. - Note this is different to EDA which is a more free-flowing narrative - instead this section must tell the reader exactly what they need to know about your dataset - not everything that you tried, or the journey you went on. - Code for these aspects does not have to be included (and likely should not be included) as part of the main content but must be included in appendices or in linked GitHub repos. - Shortcomings or weaknesses of the survey, sampling approach, and/or analysis are thoroughly discussed.								
Numbering All figures/tables/equations/etc are numbered and referred to in the text.	2.0 pts Yes			1.0 pts Partially			0.0 pts No	
Proofreading All aspects of submission are free of noticeable typos.	2.0 pts Yes			1.0 pts Partially			0.0 pts No	
Graphs, Tables, etc. Any graphs/tables/images/summary statistics that are used must be of an incredibly high standard. <input type="checkbox"/> They must be well formatted and report-ready <input type="checkbox"/> They should be clean and digestible. <input type="checkbox"/> They: 1) serving a clear purpose; 2) fully self-contained through appropriate use of labels/explanations, etc; and 3) appropriately sized and coloured (or appropriate significant figures in the case of stats). - Code for these aspects does not have to be included (and likely should not be included) as part of the main content but must be included in appendices or in a linked repo that is mentioned in your main content.	6.0 pts Exceptional	5.0 pts Great Overall great, but one aspect could use improvement.	4.0 pts Good Overall great, but two aspects could use improvement or one aspect is not done.			3.0 pts Adequate Adequate, but made more than one or two major issues.	2.0 pts Barely Adequate	1.0 pts Poor Poor, me: two or mc major issi

References Criteria - All data/software/literature/etc are appropriately noted, and cited in a consistent way that matches a major citation style. - You must cite the software and software packages that you use. - You must cite the datasets that you use. - You must cite any methodologies or ideas (not directly covered in the lectures or pre-requisite courses) that you used.	4.0 pts Exceptional	3.0 pts Acceptable	Ratings	0.0 pts Poor or Not Complete
	Consistent, complete, and well formatted.	Minor issue (in formatting, or slightly incomplete)		Severely incomplete and/or major issue
Reproducibility - Given the circumstances, the report and analysis appear to be as reproducible as possible. - Given the nature of how the GSS is obtained, you do not have to use code to actually download it but you must document how you got it. - This likely involves features such as the use of code to import data, prepare it, create plots, conduct analysis, and generate documents; the use of seeds where possible. - Code examples from class are over-documented for the purpose of helping you learn - e.g. in your submission you wouldn't need to explain what 'mutate' is doing. - If you screenshot your code and then include the image then this will be given 0 as it is not possible to test. - You must appropriately document your scripts such that someone coming in could follow them. - Your repo must be thoroughly organized.	6.0 pts Exceptional	5.0 pts Good Minor issues that do not affect the ability of the code to be run on another computer.	4.0 pts Adequate	2.0 pts Poor Substantial shortcomings, such as few comments.
General Excellence General excellence marks are available because there are always some students who clearly excel in a way that has not been anticipated. It would be unfair to not reward that excellence, but it is difficult to plan for. These general excellence marks are available to all students, and it is expected that all students will attempt to excel in some way that is not anticipated if they want to get 100 per cent.	3.0 pts Exceptional	2.0 pts Great	1.0 pts Well done	0.0 pts No Additional Points

Drop here!