

Final Project Part 2
Analysis Plan Flowchart
Due: November 28, 2021 by 11:59PM ET on Quercus

Goal of the Assessment:

The purpose of this assessment is for you to think about how you and your partner approached your analysis in the Video Project and to construct an analysis plan for your final project. By formulating a plan early, you can ensure that you perform your analysis using the correct methods at the correct times. Your detailed plan will also aid you in writing out your final report, as it will form the foundation to your Methods section, where you will outline how you plan to answer your research question and provide justification for any decisions you will make along the way. It will also let you demonstrate that you understand how to use the various aspects of linear regression, and the limitations of these tools.

Instructions:



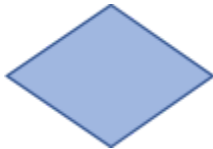


You are tasked with creating a [flowchart](#) that outlines the steps you will take in your final project analysis to arrive at an answer to your proposed research question. Since you have already completed your exploratory data analysis (EDA) in Part 1 of the project, your flowchart can start by outlining any decisions being made based on the results of your EDA and moving forward into the steps you will take to arrive at a final model that answers your research question. **You will not need to actually code anything or do what you describe in your flowchart for this assessment.** Rather focus on the results of your EDA and how you will build models and make decisions to reduce or increase the size of your model, as well as how you will incorporate model diagnostics. This means you will need to think about the order (i.e. where in the analysis it makes sense to do these steps) in which you use certain linear regression methods and what process you may need to follow/decisions you might need to make depending on the results of these methods.

The plan that you are outlining will be very general – while it should be written using the variable names and properties in your dataset, you will not necessarily know the outcome of any decision step. Rather you should consider all possible outcomes of these decisions and show us in each case what your steps would be. You should focus on making each element of your flowchart (e.g. process, decision, etc.) clear and specific (e.g. don't just say "check assumptions" but be specific about how (specific residual plots) and what you're checking (looking for patterns)), so that we know exactly what you would be doing at each point and how that leads to different decisions and further steps. If you've already identified interesting characteristics of your data, or important predictors based on reviewing the literature, you should include this at the starting point of your flowchart.

How to Create a Flowchart:

A flowchart consists of different shaped boxes in which you write steps/decisions corresponding to a process in your analysis. Arrows connect these boxes and point in the direction that you would proceed after each step. You can create loops and branches in your flowchart if you need to repeat a certain process until a desired result is achieved (loop) or if, depending on the result, you will take a different approach (branch). There are specific shapes for the boxes that have specific meanings, which you should use (see below, table 1). These boxes should contain text that describes what is occurring at this step (e.g. check assumptions by creating residual plots).

Table 1 - Flowchart Elements and their meaning/purpose

Flowchart Elements	Meaning and Function
	This shape indicates the starting point or ending point in your flowchart. You should ensure it is clearly written inside the box what the starting point of your process is and exactly where it ends.
	This shape indicates that you are applying a certain method/process in your analysis (often called an action). The text inside should specify exactly what action you are taking at this step.
	This shape indicates you must make a decision. The text inside this shape should refer to what question you are asking at this step and should have arrows leading out indicating what you would do depending on whether the answer is yes/no, for example.
	If you need to continue your chart on another page/on another part of the same page, you can use a circle to indicate that there is a break in your flowchart at the end of one page/step and use the same circle as the starting spot on the new page/area. You should include a label inside that is the same on both pages/areas so that it's obvious this links the two halves of the flowchart together.
	The arrow is what connects the steps together and shows how you will proceed through your analysis. Arrows are unidirectional but can be used to refer to a previous step (if looping is needed).

Some details and examples on these flowchart elements can be found at https://www.rff.com/flowchart_shapes.php and some examples of the types of structures you may wish to consider (e.g. how to write an iterative or branching step) can be found at https://www.rff.com/structured_flowchart.php. These should be used to help you decide how you might want to draw your flowchart (while they make reference to a software, don't purchase this). An example flowchart has been provided on the Quercus assignment page.

You may create your flowchart in any way that you are comfortable. This includes simply drawing it by hand, creating it in Word or some other word processing program, or using more graphics-based software if you have access to one. It is possible to create flowcharts in LaTeX (using the tikz package) however it is often quite cumbersome to do and therefore not encouraged unless you are already familiar with its use. You should use the shapes and flowchart elements above as they are intended to be used, and feel free to colour code your

flowchart in any way that you see fit. If you do colour code your flowchart, please provide a legend to tell us what each colour refers to.

Submission Requirements:

Your final flowchart should be submitted to Quercus by the deadline above. No late submissions will be accepted. Your submission should have the following characteristics:

- Be no more than 2 pages (of standard letter-sized paper) in length
- Should be saved as either a PDF or .docx file
- All components and text should be legible and easy to read.
- If colours are used to distinguish certain aspects of the analysis, a legend must be included.
- Your flowchart should include steps relating to material on linear regression covered throughout the course, from **Module 2 up to Module 8's material on collinearity** (the rest of module 8 does not need to be incorporated into the flowchart).

Rubric:

Quality	Excellent (3 points)	Good Effort (2 points)	Needs Improvement or Complete (1 point)	Missing or Insufficient or Incomplete (0 points)
Coverage of Course Content	At least 1 topic from each relevant module is included in the flowchart.	At least 1 topic from 6 of the relevant modules is included in the flowchart.	At least 1 topic from 4-5 of the relevant modules is included in the flowchart.	Topics from fewer than 4 of the relevant modules are included in the flowchart.
Correctness of Steps	All steps/actions and decisions are correct, occur in the correct order, AND no important analysis elements are missing.	There is at most one step/action or decision that is either incorrect, occurs in the wrong order, OR there is at most one analysis element that is missing.	There are at most 2-3 steps/actions or decisions that are either incorrect or occur in the wrong order, OR at most 2-3 analysis elements are missing.	Four or more steps/actions or decisions are either incorrect or occur in the wrong order, OR four or more analysis elements are missing.
Clarity of steps and decisions	All steps/actions and decisions are phrased	No more than 2 steps/actions or decisions are	Between 3 and 4 steps/actions or decisions are	More than 4 steps/actions or decisions are

	clearly so that it's obvious what is being done at each step.	phrased so as to make it questionable what is being done at these steps.	phrased so as to make it questionable what is being done at these steps.	phrased so as to make it questionable what is being done at these steps.
Meets Submission Requirements	-	-	Flowchart elements are used correctly AND flowchart is legible and easy to read AND flowchart is no more than 2 pages AND if using colour, a legend has been provided	Flowchart elements are not used correctly OR flowchart is not legible and easy to read OR flowchart is longer than 2 pages OR if using colour, a legend has not been provided.