Overview

Think of this project primarily as an opportunity to show off what you’ll learn about data analysis throughout the course. This is a huge component of the course (250 pts), though the basic assignment is fairly simply stated: form a group, find an interesting topic to study, design a study (with potential for formal inference), get some data, analyze those data, then write up a report and make a presentation. You might have little sense for how to go about this task, so there will be lots of “milestones” to meet which will give this assignment some structure and lots of suggestions to help you avoid common problems and how make this project work out great. Keep in mind that I CAN help your group at any stage of the process. So throughout the semester, please take advantage of my time and expertise and willingness to help.

Deliverables and Timeline

Careful, early planning can make a huge difference in the quality of your project. Experience suggests that 1) advice on how to plan and carry out a group project like this, 2) opportunities for early, frequent feedback, and 3) deadlines that help motivate the group to avoid putting off the work until it’s too late to do a decent job which makes the task easier. To these ends, here’s what your group must deliver and when it’s all due:

<table>
<thead>
<tr>
<th>Checkpoint</th>
<th>Due Date</th>
<th>Credit</th>
<th>Message Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Group roster</td>
<td>Sun, Sept 30</td>
<td>5 pts</td>
<td>“Group Roster”</td>
</tr>
<tr>
<td>2. Initial project proposals</td>
<td>Sun, Oct 7</td>
<td>20 pts</td>
<td>“Initial Proposals”</td>
</tr>
<tr>
<td>3. Revised project design</td>
<td>Tues, Oct 23</td>
<td>20 pts</td>
<td>“Revised Design”</td>
</tr>
<tr>
<td>4. Data file</td>
<td>Tues, Nov 20</td>
<td>10 pts</td>
<td>“Data File”</td>
</tr>
<tr>
<td>5. Analysis outline</td>
<td>Thu, Dec 6</td>
<td>10 pts</td>
<td>“Analysis Outline”</td>
</tr>
<tr>
<td>6. Project Summary</td>
<td>Sun, Dec 9</td>
<td>5 pts</td>
<td>“Project Summary”</td>
</tr>
<tr>
<td>7. Individual Assessments</td>
<td>Wed, Dec 11</td>
<td>10 pts</td>
<td></td>
</tr>
<tr>
<td>8. Written report</td>
<td>Wed, Dec 11</td>
<td>100 pts</td>
<td></td>
</tr>
<tr>
<td>9. Group Presentation</td>
<td>Wed, Dec 11</td>
<td>50 pts</td>
<td></td>
</tr>
<tr>
<td>10. Peer Project Critique</td>
<td>Wed, Dec 11</td>
<td>20 pts</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>250 pts</td>
<td></td>
</tr>
</tbody>
</table>

Each of these checkpoints is described in detail below. All but the final report, the presentation, and the group dynamic assessment must be delivered electronically via e-mail to my Witt account by 5:00 PM on the dates above; I’ll respond electronically as well to these intermediate checkpoints. (All but the data file should be typed in the body of an e-mail message, rather than included as an attachment. Only the data file should be attached.) Only one person from the group need submit the group’s product for each checkpoint, and that person will be responsible for disseminating my feedback to the rest of the group. (If the group’s electronic contact person lists all group members as recipients for the message, I’ll “reply to all”, making this dissemination trivial.) Please use the prescribed message subject headers
above, so that I can more easily sort through the messages I’ll get from our class pertaining to these projects throughout the semester.

**0. Assessment Criteria**

To aid in your planning and to help you make an informed choice of project ideas let me share with you up front the criteria used to evaluate your final project report:

**GENERAL:** Is the topic original, interesting, and substantial – or is it trite, pedantic, and trivial? How much creativity, initiative, and ambition did the group demonstrate? Is the basic question driving the project worth investigating, or is it obviously answerable without a data-based study?

**DESIGN:** Are the variables chosen appropriately and defined clearly, and is it clear how they were measured and/or observed? Did the group control the effects of lurking variables – using randomization, stratification/blocking, and the protocol they established either for sampling individuals or for administering the treatments and recording the data? Is there enough data to make meaningful conclusions?

**ANALYSIS:** Are the chosen analyses appropriate for the variables/relationships under investigation, and are the assumptions underlying these analyses met? Are the analyses carried out correctly? Is there an effective mix of graphical, numerical, and inferential analyses? Did the group make appropriate conclusions from the analyses, and are these conclusions justified?

**WRITING:** How effectively does the report communicate the goals, procedures, and results of the experiment? Are the claims adequately supported? How well is the report structured and organized? Are text and analyses effectively interwoven? Does the writing style enhance what the group is trying to communicate? How well edited is the report?

These are essentially the same criteria you are to use when evaluating each other’s projects on presentation day. But because you won’t get a chance to read each other’s written reports, there will be a slightly modified “Presentation” category in place of the above “Writing” category.

**1. Roster - Sun Sept 30**

Form a group of 3 or 4 students from our class. Have one person send me a group roster electronically (to my Witt e-address, bshelburne@wittenberg.edu) by the date listed above, using the message subject header “Group Roster”. Note: If I have to help you find or form your group, then you can’t get full credit for this checkpoint. So take the initiative to ask around and find a group to work with.

**2. Initial Proposals – Sun Oct 7**

Please submit two proposals. Too often a group locks their focus on a single idea – often the idea they thought of first – and doesn’t realize that it’s unworkable, or that there are many other ideas that would
work out at least as well or even better. Moreover, you have learned about two types of studies involving data, and I’d like to check whether you understand the basics of both types. So I’ll require that you submit two separate proposals: one observational study and one experimental study. Count on brainstorming on at least half a dozen serious ideas before you can groom two of them into mature proposals. I’ll respond to both of them, and I won’t hesitate to give you my advice on which proposal seems more workable for this project. As with the project rosters, please submit your proposals (in the body of an e-mail message, not as an attachment) to me electronically, using the message subject header “Initial Proposals”. I’ll try to respond with my feedback within two weekdays of this deadline.

For the most part, I leave the choice of topic up to you. But please avoid banal topics or trite ideas that really don’t work – especially anything related to alcohol consumption or other partying behaviors, shoe size, GPA, growing plants, taste testing, the effects of caffeine on pulse, testing the strength of paper towels, laundry stain removal, mold on bread or cheese, etc. You might initially gravitate toward data that you can collect on Witt students; unfortunately, these are often the least fruitful ideas and don’t work (see below). To give you incentive to think outside the Witt student box, it is required that at least one of your proposals must involve individuals that aren’t humans. Try to pick something that’s interesting yet substantial and worth studying, and aim for a topic that you think nobody has tried before; remember that part of your overall grade will be based on originality.

Note: Due to non-response problems, observational studies based on surveying Wittenberg students and/or faculty are prohibited!

Once you decide on a general topic, think hard about what kind of data you’ll gather. You’ll have fewer headaches with this project if you select variables that you can define clearly and measure or observe reliably. In particular, avoid variables that require human subjects to make estimates about behaviors that are hard to quantify or hard to keep track of. For example, “time spent studying” might sound pretty simple until you realize how tough it is to specify what constitutes “studying”, how much a person’s study patterns vary from day to day and week to week, how little most people pay attention to precisely how much time they spend doing any particular activity, etc. Select topics for which you can get some fairly reliable data on some fairly concrete variables.

Keep in mind that we’ve analyzed both quantitative and categorical variables. Thought you haven’t yet learned why, it typically takes much more data to notice interesting trends when you’re analyzing categorical variables. Unfortunately, it’s often simpler (though lazier) to think in terms of categories (e.g., small/medium/large sized cars) when there’s really some interesting quantitative variable underlying these categories (e.g., the measured length of the cars, in inches). So as you pick a topic, you’ll be doing yourselves a favor if you gravitate towards topics for which you can measure or observe mostly quantitative variables, though with a couple categorical variables thrown in for group comparisons.

Make sure you can demonstrate your mastery of the course material through this project, and that you’ll be able to pull some neat conclusions from your studies. Make sure that the overall scope and workload is comparable for all groups. To those ends use the following constraints for your project proposals:

- Have enough data to make meaningful inferences. There is no magic number of individuals required for all projects. But aim for at least 200 sampled individuals in your observational study, and make sure there are at least 20 individuals in each category of each of your
categorical variables. For your experiment, it’s not so simple. Ultimately you’ll need at least a
certain number of subjects per treatment group. But as you’ll see later in the course, you need
a lot more data to make worthwhile conclusions if your response variable is categorical (say, at
least 50 per treatment group!) than if your response variable is quantitative (say, at least 15
per treatment group). Generally, more data is better – though I realize that costs more time,
money, and effort.
• Have enough variables of both types to show that you know how to handle relationships of all
sorts. So for each proposal be sure you have at least two categorical variables and at least two
quantitative variables. Again, involving more variables is better at this stage, so that you’ll be
more likely to have some interesting relationships to investigate.

Finally for the proposal itself

1. List the group members (again)

2. For the observational study proposal:

   A. What’s the overall purpose for the study? Describe the topic/phenomenon you want to study.
      Cite some specific hypotheses or other initial suspicions that you’ll check.
   B. Specify the individuals in your study. Describe the population of all such individuals. How many
      individuals are there in the population? (If you don’t know, at least estimate this roughly.)
   C. How you will sample the individuals you’ll include in your study? How many individuals will you
      sample?
   D. Specify all the variables that you’ll measure or observe on each individual – including their type
      (quantitative or categorical), how they’re defined, and how they’re measured or observed. For
      categorical variables, specify the possible categories. For quantitative variables, give units of
      measurement (when appropriate).
   E. Is one (or more) of the variables a “response” variable? If so, which one(s)?

2. And for the experiment proposal:

   A. What’s the overall purpose for the study? Describe the topic/phenomenon you want to study.
      Cite some specific hypotheses or other initial suspicions that you’ll check.
   B. What are the experimental units/subjects/individuals? How many will you use? How will you
      select individuals for your study? (Note: this is not where randomization is absolutely required
      in an experiment – though you’re welcome to use randomization at this stage if you want to
      generalize to a larger group of individuals.)
   C. List each factor (i.e., explanatory variable under your direct control), and specify whether each is
      quantitative or categorical. How many levels will you use for each factor? (Keep the number of
      levels small and manageable!) List the treatments.
   D. What is the response variable? How is this variable defined, and how will you measure or
      observe it on each individual?
   E. How many individuals will be assigned to each treatment group? How many individuals are then
      required for the whole experiment?
   F. Describe any blocking variables you’ll use. List other variables that could be related to your
      response variable, but which will not serve as factors or blocking variables. (You should
      measure or observe these variables on each individual anyway, just so you can later show that
      randomization helped balance the effects of these variables across the treatment groups.)
3. Revised Design – Tues Oct 23

Once I respond to your proposals, I’d like you to choose one of your ideas (or scrap them both and submit a new proposal), then refine this idea in light of my feedback. You should supply essentially the same information required for the initial proposal, but give a bit more detail. In particular...

If you choose your observational study:

1. Describe the general topic/phenomenon you want to study, as well some focused questions that you hope to answer or some specific hypotheses that you intend to assess.
2. Specify what the individuals are, describe the larger population/phenomenon to which you’ll try to generalize, and (if appropriate) estimate roughly how many such individuals there are in the population.
3. Describe in detail how you’ll go about selecting individuals from this population for use in your study: what you will use as a frame, whether you will stratify on any variable(s), how you will use randomization, etc. How many individuals will you include in your study?
4. Describe the variables that you’ll measure or observe on each individual: carefully define each variable and describe how each will be observed or measured. For categorical variables, list the possible categories; for quantitative variables, specify the units of measurement.

If you choose your experiment:

1. Describe the general topic/phenomenon you want to study, as well some focused questions that you hope to answer or some specific hypotheses that you intend to assess.
2. Specify what the experimental subjects/units are. Describe how you’ll select these experimental units for use in your experiment. (This need not be through randomization, unless you want to generalize the results of your experiment to a larger group.)
3. Cite the response variable. Is it quantitative or categorical? How is it defined, and how will you measure/observe it?
4. List the experimental factors, and cite the number of levels of each factor. List the treatments (i.e., the combinations of factor levels).
5. Describe the protocol for administering the treatments to the experimental units.
6. Will there be a blocking variable? If so, name the variable and list the blocks.
7. Describe how you’ll assign experimental units to the treatment groups. How many experimental units will be assigned to each treatment, and how many are then required for the entire experiment?
8. List other possibly related variables (not under your control) that you’ll measure or observe for later analysis.

Submit this final proposal electronically (in the body of an e-mail message, not as an attachment) by the due date above, using the message subject header “Revised Design”, and I’ll try to respond within two of our class days, so that you can then get rolling with your data collection.
4. Data File – Tues Nov 20

There are three motivations for this checkpoint:

1) I need a copy of your data – to check your analyses, but also to use your data in future versions of this class.
2) I want to make sure you keep moving forward on this project and don’t put it off.
3) I’d appreciate an extra opportunity to give you some advice.

Send me your data by the deadline above, within the following constraints – most of which are designed to make your data file understandable and hence useable to anybody who’s not directly involved in your project (viz., to me and to future students):

- The data must be in a Minitab file (not Excel or some other format).
- Save your data as a Minitab “Worksheet” – not a Minitab “Project” file. Here’s how: File > Save Current Worksheet As, and make sure the “Save as type:” field is simply “Minitab”, which will save the file as a worksheet, as desired.
- Give the file a descriptive name that clearly communicates the context and distinguishes it from the other groups’ data files, e.g., “Quality and Cost of US Colleges”, certainly not “Data” or even “Our 127 Statistics Project Data File.” In fact, I absolutely forbid you from using the following words in your file name: project, data, file, Minitab, statistics, worksheet.
- Keep in mind the distinction between individuals and variables. Each individual should be on a separate row of the data file, and each variable should be in a separate column.
- Include a column that identifies the individuals; use numerical labels if the individuals must be kept anonymous. (If you don’t have any names or labels for the individuals, create an ID column and simply number them 1, 2, 3, …. Here’s how to do this quickly: Calc > Make Patterned Data > Simple Set of Numbers.)
- Name all variables helpfully and contextually, e.g., use “Airport” and “Water Temp”, not “Individuals” and “Treatments”, and certainly not “A” and “B”.
- Similarly, for the category names, use whole words and phrases, not cryptic codes, e.g., use “Male” and “Female”, not “1” and “2”, or even “M” and “F”.
- That said, try to limit your variable and category names to about a dozen characters, or else Minitab will mercilessly truncate these names when labeling reports and graphs. This may take some abbreviation.
- Avoid the mistake of naming a variable by essentially listing its categories, e.g., use “Sex”, not “Male/Female”.
- Also avoid the mistake of naming a variable for its units of measurement, e.g., use “Age” and “Study Time”, not “Years Old” or “Hours Studying”.
- **Be sure that are sufficient numbers of individuals in each category of each categorical variable!** If there are categories with too few individuals for you to spot any trends or to make meaningful inferences, create an additional version of this variable with fewer, consolidated categories (perhaps including an “Other” or “Miscellaneous” category).
- Order the variables logically, too (highlight one or more columns, then select Editor > Move Columns), with the column identifying the individuals first. If you have a response variable in your study, that typically goes second.
- Adjust the display format of each variable (click anywhere on the column, then select Editor > Format Column) so that the fields have an appropriate width.
• For categorical variables that have a natural ordering, tell Minitab how the categories should be ordered: click anywhere on the column, then Editor > Column > Value Order.

• Be sure that Minitab recognizes each variable’s intended type by checking the column number: “C3” indicates that Minitab can treat it as a quantitative variable, but “C3-T” indicates a text (categorical) variable. Minitab cannot do any of the analyses appropriate for quantitative variables if there are any non-numeric characters in anywhere in the cells of the column. Use Data > Change Data Type to fix any problems.

• Check for typos! Manual inspection is OK, but it’s tedious and it’s easy to overlook misspellings. Running some simple analyses can more quickly make most data entry errors obvious. Try Stat > Tables > Tally for categorical variables, and a dotplot for quantitative variables.

• Check for “ghost” entries and variables – where spaces or something not visible was inadvertently entered in various cells. Here’s how: look at the Project Manager, which is that little window that’s normally minimized but is always lurking in the background. (If you can’t find it, try Window > Project Manager.) Click on Columns, which then displays all the variables that Minitab thinks are in your worksheet, along with their type and the number of observations in each of them. Delete any unnecessary variables or rows – with Data > Erase Variables and Data > Delete Rows.

Even though you’re required to end up with longer, descriptive variable and category names, please take advantage of Minitab’s features to speed up the data entry: autofilling cells, entering short codes and then replacing them with longer descriptors, and using Calc > Make Patterned Data.

By the due date, send me your group’s file as an attachment to an e-mail message with the subject header “Data File”.

5. Analysis Outline – Thu Dec 6 (Fri Dec 7)

At this point you should have your data in hand. Think about how you’ll structure and organize the analysis of your data. I’m interested partly in whether you’re following our principles of data analysis. (This would be a good time to review those principles.) I’m also interested in whether you’ve chosen appropriate tools for the types of variables you’re analyzing; you should know that the tools appropriate for any given analysis will depend largely on which types of variables are involved. But of course I’m interested in whether you’re focusing on analyses that can give you some insight into the main question/motivation driving your project.

Here’s the sort of information required for this checkpoint: In some kind of outline form,

- List the order in which you’ll analyze the variables and their relationships. (Remember to analyze each variable separately before analyzing it in relationship with other variables.)
- Under each distribution or relationship, list the tools that you’ll use to analyze that distribution or relationship. Within each analysis, move from visual tools to numerical tools to (if appropriate) inferential tools.
For example:

**Distribution of Assault Type (categorical)**
- **visual:** bar chart
- **numerical:** frequency table
- **inferential:** confidence interval for proportion of assaults involving a gun

**Relationship between Assault Type (categorical) and Age of Assailant (quantitative)**
- **visual:** parallel dotplot
- **numerical:** descriptive stat’s for Age, broken down by Assault Type
- **inferential:** F-test for difference among mean Ages, across Assault Types

The best strategy for organizing the analysis will depend on the particulars of your design and your data, of course. But I can give you some more general advice:

- For observational studies, it may be helpful to start with a very quick series of single-variable analyses that help characterize the variety of individuals involved in your study.
- If your study is an experiment, you should check whether your use of randomization was effective – i.e., whether the treatment groups started out balanced with respect to any potential lurking variables that you also measured.
- If your study has a response variable, you may want to focus on the distribution of this response first, and then mostly on the relationship between each explanatory variable and the response variable.
- You definitely need not include an analysis of every relationship between every pair of variables, especially if your group has more than the minimum number of variables required. In fact, you’ll probably have time for only about 4-6 relationships in your presentation. You can carry out more analyses in the report, though you need not do so.
- Nor is it absolutely necessary to carry out formal inferences for every distribution or relationship that you analyze. This is especially true for the distribution of variables separately – unless you want to estimate a particular mean or proportion with a confidence interval, or unless you had in advance a specific value of such a parameter to test.
- Let the order of the variables and relationships dictate the overall structure of your analysis, rather than the types of tools used. In other words, don’t have a section of all of your graphs, then a section of all of your numerical summaries, and then a section of your formal inferences.

As always, please get some advice from me early and often on this crucial step. As with writing any paper or report, the effort you spend organizing at the beginning can save you from wasting all sorts of effort later, and is usually the biggest factor in the quality of the final product. So the more thought you can put into this stage, the easier it will be for you later.

As with almost all of the earlier checkpoints, submit this in the body of an e-message (not as an attachment, please) to me by the due date, this time with the subject header “Analysis Outline”.
6. Abstract/Summary – Sun Dec 9

Think of this as an opportunity to communicate the goals and results of your experiment to your peers. Focus not merely on the questions that your project is designed to address, but on the actual answers to those questions – e.g., not merely “We ran an ANOVA F-test to see whether airport size and region were related,” but “Airports in the central part of the country were much larger than those on either coast.” I’ll post the abstract/summary for each group on our web resource or via e-mail. My main motivation for doing this is to help you get acquainted with each other’s projects prior to presentation day, so that you won’t have to base your peer critiques on those short presentations alone, and so that you can prepare questions to ask each group. Please limit this to 200-300 words.

Presentation Order Preference

At this time include your group’s preference for when you would like present: first, early, middle, late, last or don’t care. I will try to reconcile them and order the Abstract/Summary for each group in the order of presentation.


Ideally, all group members would be equally involved and able and committed to the project. In reality, it rarely works that way. I’d like to reward people fairly for their efforts in this group endeavor, because it’s inevitable that there will be variation in how high a priority people put on this class and how much effort they put into this project.

To this end, I’d like each of you (individually) to describe for me how well (or how poorly!) your project group worked together and shared the load. Also give some specific comments describing each member’s overall effort. Were there certain group members who really put out exceptional effort and deserve special recognition? Conversely, were there group members who really weren’t carrying their own weight? And then, at the end of your assessment, estimate the percentage of the total amount of work/effort done by each member. (Be sure your percentages sum to 100%)! For example, suppose you have 4 group members: A, B, C, and D. In the (unlikely) event that each member contributed equally, or you could assign

25% for member A
25% for member B
25% for member C
25% for member D

Or in case person C did twice as much work as each other member, you could assign

20% for member A
20% for member B
40% for member C
20% for member D
Or if member B didn’t really do squat, you could assign

- 32% for member A
- 4% for member B
- 32% for member C
- 32% for member D

I’ll find a fair way to synthesize the (possibly conflicting) assessments within each group incorporating this assessment of effort and cooperation in each individual’s overall grade. Don’t pressure one another to give everyone glowing reports unless it’s warranted, and don’t feel pressured to share your reports with one another. Just be fair to yourselves and to one another.

Does this make sense? Let me know if you have any questions.

Because the work for this project will likely continue until the last minute, the Group Dynamic Report will be due at the beginning the final presentation time slot. Please turn in your assessment at that time, on a single word-processed sheet of paper.

8. Report – Due Tues Dec 11

First of all, your group will submit a single report; there’s no need for each group member to submit a separate report. I realize that you probably haven’t ever written a formal data analysis report before, so let me propose the following five part structure for this report:

- **Title, Date, Authors, Abstract:** Begin your report with a title page, bearing nothing except the title, date, authors, an abstract (i.e., a short, pithy paragraph summarizing the topic and results of your entire project), and the honor pledge (with each of your signatures). This abstract should be about 100-200 words long.
- **Introduction:** Begin the body of the report with a short section introducing the topic and describing why it’s of interest. Include your initial suspicions about what you thought you would learn from your data.
- **Overview of Design:** Continue with another section describing the design of your study (including a description of the individuals/experimental units and variables), much as you did in your initial proposal, though in more detail.
- **Data Analysis:** Next comes the main section – the data analysis itself. This part will be longer by far than everything else combined; break it up into as many subsections as you see fit, **keeping in mind our data analysis principles as you organize and present your work.** Refer to your analysis outline as you structure this major section.
- **Summary:** Finish with a section summarizing the results. Also critique your own study, citing its good and bad points, and suggest how you could’ve improved it. Develop several follow-up questions for future analyses that are stimulated by your study and your results.

The document should be written in clear, concise, correct English. Just as with any formal writing assignment, mechanical mistakes and bad stylistic habits distract the reader from the points you’re trying to make. Precise use of our specialized terms is essential for effective technical communication. Aim for the usual college-paper audience – an intelligent person who doesn’t necessarily know a lot
about your subject – though you can assume that the reader knows as much about statistical methods and ideas as you’ve learned in our course.

Here’s how much weight will be put on each of the criteria, described earlier:

- 15%: General
- 25%: Design
- 45%: Analysis
- 15%: Writing

Submitted copies of the report will be kept for my records, so be sure to keep your own copy of the final product. Don’t bother with a fancy binder; a simple stapled copy is preferred.

For the most part, you’ll do well as long as you keep in mind the principles of Exploratory Data Analysis. I do, however, want to remind you to give any suspected outliers a royal treatment: find out the story behind each one, and check the influence on the various analyses when you omit these observations. Also, try to imagine how your analyses are affected by the way you measured/observed your variables, or by the way you selected your individuals/experimental units. Be sure to cite possible “lurking” variables that could help explain why the distributions or relationships turned out as they did.

How much output should you include in your report? Where should it go? Good questions. Here’s my best general advice: **Focus on your own discussion and interpretation**, and use the software’s plots and calculations primarily to back up your own claims and analysis. So at the very most, include output only if you discuss and analyze it in your text. Avoid abusing important jargon (e.g. “random”, “correlation”) which have very precise technical meanings.

Where should you put the software’s results? Optimally you should import the relevant output into the appropriate spot of your document, just before or just after the discussion. This is easy for text-based output, which you can simply copy and paste directly from Minitab’s session window to a word processing document. For plots it takes an additional step: when you’ve dressed up a plot with our software and are satisfied with it, right-click anywhere on the plot and select Copy Graph, and then you can simply paste it into most word processors, exactly in the part of your report where you’re discussing it. The next best option for the position of a plot in your document would be to insert it on a separate page right after you refer to it in your report. Please don’t simply append a stack of output pages to the end of the report. Regardless of where you include output, be sure to trim away all irrelevant detail, and include only what your reader really needs to see. And remember to label all aspects of the plots appropriately.

Let me rephrase my best general bit of advice: Don’t assume that the computer’s output will speak for you! **Focus instead on your own comments, interpretation, and explanation** of the plots and numbers that the software spits out. And make sure you move beyond merely reporting numbers (which is usually unnecessary anyway if you present the relevant output) to interpretation in context. In case I haven’t hit you with this enough yet this semester, here’s another example:

Suppose you’re reporting measures of center for the distance covered by recently retired Witt math prof Eric Wilson on the 68 days of his cross-country bike trip, broken down by region of the country. Here are several ways to write up this part of the analysis, from worst to best:
1) Least effective approach: Merely import undigested computer output:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Region</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>East</td>
<td>17</td>
<td>63.71</td>
<td>62.50</td>
<td>63.55</td>
<td>12.35</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>18</td>
<td>69.11</td>
<td>68.35</td>
<td>69.63</td>
<td>17.75</td>
</tr>
<tr>
<td></td>
<td>West</td>
<td>33</td>
<td>58.11</td>
<td>59.40</td>
<td>58.80</td>
<td>11.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Region</th>
<th>SE Mean</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance</td>
<td>East</td>
<td>3.00</td>
<td>43.80</td>
<td>86.00</td>
<td>56.00</td>
<td>70.85</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>4.18</td>
<td>29.70</td>
<td>100.10</td>
<td>61.68</td>
<td>80.52</td>
</tr>
<tr>
<td></td>
<td>West</td>
<td>1.97</td>
<td>20.00</td>
<td>79.70</td>
<td>53.80</td>
<td>63.85</td>
</tr>
</tbody>
</table>

So far the computer has done all of the work and you’ve done almost nothing.

2) Slightly better: Trim and edit the output:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance Covered in East</td>
<td>63.71</td>
<td>62.50</td>
</tr>
<tr>
<td>Middle</td>
<td>69.11</td>
<td>68.35</td>
</tr>
<tr>
<td>West</td>
<td>58.11</td>
<td>59.40</td>
</tr>
</tbody>
</table>

At least you’ve now eliminated some of the clutter and have focused on a comparison of our usual measures of center. But you still haven’t shown any thoughts of your own.

3) Also does not work simply reporting raw numbers in your text:

The means were 63.71 for the east, 69.11 in the mid, and 58.11 in the west. The medians were 62.50 in the east, 68.35 in the mid, and 59.40 in the west.

This is just a text-based version of the table above. What do these numbers tell you? That’s what’s missing, and that’s what’s important.

4) Getting warmer: Make some comparisons, but not in context:

The biggest means and medians are in the mid, followed by east, followed by west:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance Covered in East</td>
<td>63.71</td>
<td>62.50</td>
</tr>
<tr>
<td>Middle</td>
<td>69.11</td>
<td>68.35</td>
</tr>
<tr>
<td>West</td>
<td>58.11</td>
<td>59.40</td>
</tr>
</tbody>
</table>
5) Better still: Interpret the comparisons in context:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance Covered in East</td>
<td>63.71</td>
<td>62.50</td>
</tr>
<tr>
<td>Middle</td>
<td>69.11</td>
<td>68.35</td>
</tr>
<tr>
<td>West</td>
<td>58.11</td>
<td>59.40</td>
</tr>
</tbody>
</table>

Typically Eric rode the farthest per day in the middle of the country, next farthest in the eastern region, and least far in the west:

6) Best: Comparisons in context, with interpretation and explanation:

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance Covered in East</td>
<td>63.71</td>
<td>62.50</td>
</tr>
<tr>
<td>Middle</td>
<td>69.11</td>
<td>68.35</td>
</tr>
<tr>
<td>West</td>
<td>58.11</td>
<td>59.40</td>
</tr>
</tbody>
</table>

Typically Eric rode the farthest per day (nearly 70 miles) in the flat midwestern states and the plains in the middle of the country, far less (about 5-6 miles less per day) in the mild river valleys and the Appalachian mountains in the east, and least (an additional 3-5 miles less per day) chugging up and down the rugged Rockies out west:

Does that make sense? The output should be there only to reinforce your own analysis. You should do this for all aspects of all distributions and relationships.

9. Presentation – Due Tues Dec 11

As many or as few group members as desired can take part in the presentation. Feel free to use any medium you want. You may bring in posters, distribute handouts, use the board, bring in transparencies to use with a projector, use PowerPoint, etc. And of course you may use our software; to that end, you can use the instructor’s computer, which is hooked up to the projector, just as I’ve done every day in class. If you need any equipment not already in our classroom, please let me know at least a couple days in advance, or bring your own equipment.

To keep the session manageable, I’d like to limit each presentation time to 12 minutes, plus up to 3 minutes for questions. A word of advice: rehearse your presentation several times, and time your presentation during each rehearsal. 12 minutes can go by very quickly, and in the interests of time I won’t hesitate to cut you off, even if you’ve presented only half of what you intended. Be sure to plan every aspect of your presentation, to minimize the chance for unpleasant surprises, and to make sure your presentation is seamless.

One more thing: no bribing your classmates with candy or other swag to get better ratings ;)

13
Presentation Day – Tues Dec 11

I’d like to make sure you all understand what will happen during our final exam time. In short: project presentations and critiques. After each presentation, we’ll leave a couple minutes for everyone to scratch down some notes while the next project group sets up. There should be 6-8 groups, so we should be finished with the presentations in an hour and a half to two hours. After all the presentations, you’ll have the rest of the exam period to write up a critique of the projects, including your own. I’ll ask you to write up your comments in a blue book, and to give each project a numerical rating (out of 20 points) on each of the four assessment criteria, described below. I’ll also give you a grid on which to report your numerical ratings.

Component Scores

Here’s a breakdown of how you’ll be assessed on the Presentation Day activities:

A score of up to 20 points for the quality of your peer critiques:

- 5 pts: How thorough, detailed, and specific are your comments?
- 5 pts: How accurate are your comments? Do your ratings reflect these comments?
- 5 pts: How consistent are your ratings of each group with the average peer ratings?
- 5 pts: How consistent are your ratings of each group with my ratings?

A score of up to 50 points, based on the presentation:

- 25 pts: for the average peer rating of the group’s project, and
- 25 pts: for my rating of the group’s presentation.

A score of up to 100 points for the written project report, using the criteria described earlier.

These guidelines are meant to be helpful hopefully providing structure so you won’t feel lost. I’m also reemphasizing that this assignment (and this course!) is about insight, not numbers; concepts, not calculations. Focus on moving beyond characterization and description, to interpretation and explanation. Context is crucial! (Where have you heard that before?)

Please feel free to get in touch with me at any point in the process if you want some clarification or advice about any of this.
## Deliverables and Timeline (Review)

<table>
<thead>
<tr>
<th>Checkpoint</th>
<th>Due Date</th>
<th>Credit</th>
<th>Message Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Group roster</td>
<td>Sun, Sept 30</td>
<td>5 pts</td>
<td>“Group Roster”</td>
</tr>
<tr>
<td>12. Initial project proposals</td>
<td>Sun, Oct 7</td>
<td>20 pts</td>
<td>“Initial Proposals”</td>
</tr>
<tr>
<td>13. Revised project design</td>
<td>Tues, Oct 23</td>
<td>20 pts</td>
<td>“Revised Design”</td>
</tr>
<tr>
<td>14. Data file</td>
<td>Tues, Nov 20</td>
<td>10 pts</td>
<td>“Data File”</td>
</tr>
<tr>
<td>15. Analysis outline</td>
<td>Thu, Dec 6</td>
<td>10 pts</td>
<td>“Analysis Outline”</td>
</tr>
<tr>
<td>16. Project Summary</td>
<td>Sun, Dec 9</td>
<td>5 pts</td>
<td>“Project Summary”</td>
</tr>
<tr>
<td>17. Individual Assessments</td>
<td>Tues, Dec 11</td>
<td>10 pts</td>
<td></td>
</tr>
<tr>
<td>18. Written report</td>
<td>Tues, Dec 11</td>
<td>100 pts</td>
<td></td>
</tr>
<tr>
<td>19. Group Presentation</td>
<td>Tues, Dec 11</td>
<td>50 pts</td>
<td></td>
</tr>
<tr>
<td>20. Peer Project Critique</td>
<td>Tues, Dec 11</td>
<td>20 pts</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>250 pts</td>
<td></td>
</tr>
</tbody>
</table>
Peer Assessment of Projects

Below are the criteria that will be used to assess each other’s projects. Note that they’re essentially identical to the criteria used to assess the reports, except that the final criterion, “Writing”, has been modified and renamed “Presentation”.

**GENERAL**: Is the topic original, interesting, and substantial – or is it trite, pedantic, and trivial? How much creativity, initiative, and ambition did the group demonstrate? Is the basic question driving the project worth investigating, or is it obviously answerable without a data-based study?

**DESIGN**: Are the variables chosen appropriately and defined clearly, and is it clear how they were measured/observed? Did the group control the effects of lurking variables – using randomization, stratification/blocking, and the protocol they established for sampling individuals or administering the treatments and recording the data? Is there enough data to make meaningful conclusions?

**ANALYSIS**: Are the chosen analyses appropriate for the variables/relationships under investigation, and are the assumptions underlying these analyses met? Are the analyses carried out correctly? Is there an effective mix of graphical, numerical, and inferential analyses? Did the group make appropriate conclusions from the analyses, and are these conclusions justified?

**PRESENTATION**: How effectively did the group report communicate the goals, procedures, and results of the study? Was the presentation polished and rehearsed? Did the group make efficient use of the available time and effective use of their chosen media?

In your blue book, write your comments about each group (including your own!) in the order in which the groups made their presentations (begin each set of comments on a new page). In the grid below, give whole-number scores from 0 to 10 on each criterion for each of the groups (including your own!), and add up the total score for each group. When you’re finished, turn in both your blue book and this rating sheet.

<table>
<thead>
<tr>
<th>Group</th>
<th>General</th>
<th>Design</th>
<th>Analysis</th>
<th>Presentation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>