**UMUC Data 620 Assignment 9.1**

Locate your very favorite multidimensional time series and tell us all about it.

**Define your Time Series**

A time series is a series of data benchmarked to a particular point in time; for example, the closing price of the stock of XYZ Corporation on December 1, 2, 3, 4, and 5 would comprise a time series of five data points.

Your time series should have the following characteristics:

1. It needs to span at least ten to twenty roughly equivalent time periods. A period may be a day, a week, a year, or some other unit of measure which makes sense for what you are measuring.
2. It needs to leave room for you to make a prediction for some time period in the future, based on the available data in your time series. For example, if you choose the years 1910, 1911, 1912, 1913, and 1914 to measure cigarette smoking behavior in Americans, your future prediction must be after this year, so you must now predict behavior for next year over a data gap which spans almost 100 years. A more sensible choice of years for this would be 1930, 1950, 1970, 1990, and 2010, which leaves you plenty of room to predict for 2030.
3. It needs to have at least three other variables for your data. These need to have occurred in the same time frames, but may be from different sources. In the example below, our main time series is the closing price of the stock, and the three additional variables were taken from other data sources.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Year | Stock closing price | Variable 1 – number of employees | Variable 2 – annual rainfall | Variable 3 – number of patents granted |
| 1 | $10 | 50 | 5 “ | 0 |
| 2 | $15 | 100 | 7 “ | 0 |
| 3 | $13 | 150 | 5.5” | 10 |
| 4 | $21 | 103 | 8” | 11 |
| 5 | $25 | 115 | 5” | 15 |

You are welcome to use any data source you choose. Some particularly useful ones can include:

* Stock price of your favorite company (finance.google.com is a good start)
* Population estimates of cities or countries
* Google Trends searches on keywords (<https://support.google.com/trends/answer/4365538?hl=en>)
* Health measures of items like the flu or longevity (the Center for Disease Control has data)
* Some data set from your own company or life (in the past I had one student chronicle a 300 pound weight loss after bariatric surgery).
* Jen Underwood has compiled a great list of publicly available data sets at <http://www.jenunderwood.com/2016/01/14/my-favorite-public-data-sources/> (scroll halfway down, to “So Many More To Explore.”
* The American Time Use Survey (ATUS) from the Bureau of Labor Statistics has fascinating time series available for just about any activity you could like, from watching TV to working. You can pull easily from a few top categories, or select your own at <https://www.bls.gov/tus/labstattips.htm#FF> (scroll down till you get to Data Extraction Tools / Top Picks.)

The following data sets are from Professor Steve Knode’s DATA 640 class, but they may be useful for you here as well:

* MovieLens <https://old.datahub.io/dataset/movielens> . This data set contains 10000054 ratings and 95580 tags applied to 10681 movies by 71567 users of the online movie recommender service MovieLens.
* Reddit’s List of Data Sets <https://www.reddit.com/r/datasets/> (this can get pretty wild!)
* Tableau’s very own List of Public Data Sets at <https://public.tableau.com/s/blog/2013/08/data-data-everywhere>
* KD Nuggets’ List of Governmental Data Sets <https://www.kdnuggets.com/datasets/government-local-public.html>
* And last, but not least, Awesome Public Datasets <https://github.com/awesomedata/awesome-public-datasets> (yes, that’s really its name … awesome!)

You will be sharing this presentation with your classmates (and possibly with the Internet at large, because you will be posting on YouTube or similarly public site. This can be excellent job exposure; several previous students have linked their LinkedIn accounts to their presentations as a sort of online resume).

You will also be commenting on some of your classmates’ presentations.

**Make Your Forecast**

* Review Tableau’s training material on trend lines (online help: <http://www.tableau.com/learn/tutorials/on-demand/trend-lines>)
* Review the training material on forecasting as well (online help: <http://www.tableau.com/learn/tutorials/on-demand/forecasting>)
* I recommend you just use their default forecast estimates – don’t bother trying to manually set any options. You are welcome to use whatever type of forecast (linear, log, exponential, seasonal) feels right to you for your data set. Note the forecast type in your presentation – i.e. “I used a linear trendline…”
* Make sure to include confidence intervals (the little funnel-looking things out to the right.) They are called ‘estimates’ in the forecasting methods video above.
* Do not use Excel, R, Python, or any other software to forecast. Just use Tableau.
* State your personal confidence in the forecast. It’s perfectly acceptable to state something like “Tableau is forecasting a value of 100 for July. I have low confidence in this forecast because {state your reasons here.}”

**Create your Story**

Determine what the story in your data is. A good story has

* A narrative, with a research question at the beginning, and a conclusion at the end. Examples of good narratives include, “Is the US crime rate really increasing?” or “What is happening to the crab population in Maryland?”
* Each graphic will support some aspect of your story.
* Each aspect of the story will be one brief caption or sentence. This will go on the top of your Story Point for your presentation.
* A good test of your narrative is to see if you can give it to somebody who is not taking this class. Do they understand what you are claiming?

Key questions to consider as you craft your story:

* Background of the data set – what motivated you to study this?
* Description of the time series and trends you see
* Any interactions or explanations from background factors? For example, if you are doing CO2 emissions, can you include average global temperatures to support or debunk global warming?
* Use annotations to elaborate on your story. (Online help: <http://onlinehelp.tableau.com/current/pro/desktop/en-us/help.htm#annotations_annotations_add.html>)

Some helpful graph types for specific data sets types:

* Bollinger bands – useful for stock price data (online help: <http://www.tableau.com/learn/tutorials/on-demand/bollinger-bands>)
* Control charts – useful if you have a time series which should stay constant over time but appears to be moving up or down, such as in a global warming context (online help: <http://www.tableau.com/learn/tutorials/on-demand/control-charts>)
* Waterfall charts – useful if you have something like cash flow or inventory, which goes up or down over each time period by a certain amount (for example, yesterday’s inventory + receipts – sales = today’s inventory). (online help: <http://www.tableau.com/learn/tutorials/on-demand/waterfall-charts>)

**Create your Presentation**

* Using your research, create a visually appealing [Tableau Story Points](http://www.tableau.com/learn/tutorials/on-demand/story-points) presentation outlining your findings.
* Make sure to include at least three time series graphs. A time series graph has the time (such as years) on the x-axis, and the data you are graphing (such as world population) on the y-axis. Usually these are line charts.
* Include up to five graphs from Tableau best representing the research results. I will view Graphs 1 through 5, and Graph #6 will not be read or graded. You may have more than five slides (for example, to include your introduction and bibliographic references). Just no more than five graphs.
* Each graph should have a reference number – for example, “Figure 1: {what you want to say about Figure 1}”
* If you include three time series graphs, that leaves room for two other graphs (say, a pie chart and a bar chart.) That’s a total of five graphs.
* Duration is up to 5 minutes long. (I will stop watching after 5 minutes.)
* This video needs to be a screen capture of yourself talking us through your Story Points presentation. You can use any screen capture software you like; if you need some, you can download a free trial of [Camtasia from TechSmith](https://www.techsmith.com/camtasia.html) or [Screen Cast-o-Matic](https://screencast-o-matic.com/) to create this video. Make sure to check the expiration date on your free trial. Be sure to use a microphone for good audio quality.
* This video should NOT be
  + PowerPoint
  + a hand-held video of your computer screen, shot using your phone
  + anything shot using the webcam of your computer
* Your presentation should start with an introduction, which should contain
  + Your name
  + The presentation title
  + This class name and number and section (for example, “DATA 620 9040, Fall 2016”)
  + Your professor’s name and the institution name (for example, “Professor Carrie Beam, UMUC.”)
  + You can do this introduction using a separate slide you make through your video production software. You can also do this using one large text box within a Tableau slide.
* Your presentation should end with a references page, which should contain
  + A citation to the source of your data (with weblink if applicable)

**Write a One-paragraph Summary**

* Write a one-paragraph (250 words or fewer) summary of your presentation.
* It needs to be an Executive Summary: if a busy reader only reads the first sentence, s/he should know what your results are. Don’t say “We analyze trends of Beanie Baby purchases.” Say instead “Beanie Baby purchases are trending upwards sharply, and we forecast sales of XXX for next quarter.”
* The best summaries contain a hook. Why should the reader continue to read the paragraph? After reading the paragraph, why should the reader want to view your video?

**Share your Video**

* Upload your video it to a publicly available website, such as YouTube or Vimeo. Make sure it is set so the link will work for your classmates and for me.

**Submit your Deliverables**

* You have two deliverables for this assignment (your one-paragraph summary and your video). You will submit both deliverables in two places.
* First, submit both your one-paragraph summary writeup and the link to your video to the Assignments Folder under Week 9.
* Second, submit both your one-paragraph summary writeup and the link to your video to our class Discussion Forum for this assignment.
* When you submit your one-paragraph writeup, be sure to paste its text into the submission box. Do not attach a Word or .pdf document that we need to click on to open before we can read it.

**Review Classmates’ Presentations**

* When you have submitted your video, review two presentations from classmates. Do try to spread your feedback around a bit; if somebody already has two reviews, try to review a presentation which doesn’t have much feedback yet.

Grading Rubric:

|  |  |
| --- | --- |
| Score | Definition |
| 90 – 100 | Exceeds Expectations. These types of presentations are really exceptional and a pleasure to watch. They meet all the criteria of the “meets expectations” presentation, and go beyond in some meaningful way. The “beyond” may be in using graphs carefully tailored to the data (stacked bar charts, or Pareto charts, for example), or it may be in the additional background research to provide informative multidimensional aspects to the time series (for example, adding population research data to a time series showing demand for electricity over time). It may also cover significantly more than 20 time periods. Storyline is clear and easy to follow. Demonstrates mastery of the visualization techniques we have covered in class. |
| 80 – 89 | Meets Expectations. The average score at UMUC is 85%, and a presentation which meets expectations will earn this score. This type of presentation will typically be easy to follow, with good audio and visual quality. It will contain between 2 and 4 time series graphs of the data, with a clear story outlined. It will also contain some graphs of other aspects of the time series, such as bubble or other plots, which illuminate a more interesting back story in the data. Interactions or background factors will be noted and explained. Each graph is labeled with a Figure number, and all APA formatting is met. A listener will come away from the presentation able to summarize the story over time. |
| 0 – 79 | Does Not Meet Expectations. Time series is weakly researched or not really a time series. This type of presentation may contain only 1 time series graph, and the balance of the graphs will be very basic (such as bar charts). The graphs may not be the best way to display the data. There is not a strong storyline, and at the end, the listener may not have a clear idea what the presentation was about.  These types of presentations also may have significant technical difficulties with sound or visual quality, or have difficulty conforming to the time limits. Figures may not be labeled, and there is divergence from APA formatting. |