Assignment 3 - Time Series, Treemaps and Multivariate Data

Part 1: Time series (9 points)

Task 1: Time Plots of Average Price

- Open *PC Universe - 2008 Point of Sale Data.txt*

A time plot is a line chart with time in the Columns shelf (x-axis).

- Create a time plot of the average weekly retail price

Note: It pays off to make sure to create a date column that is recognized as a date (as we did in Assignment 2), since the software provides many ways to explore the data at different levels of aggregation e.g. month, week, day, day of week, etc.

- Note the single null value indicated by the warning at the bottom right. Click on the warning "1 null" and filter this point out by choosing "Filter Data". Note that this adds a filter in the Filters shelf.
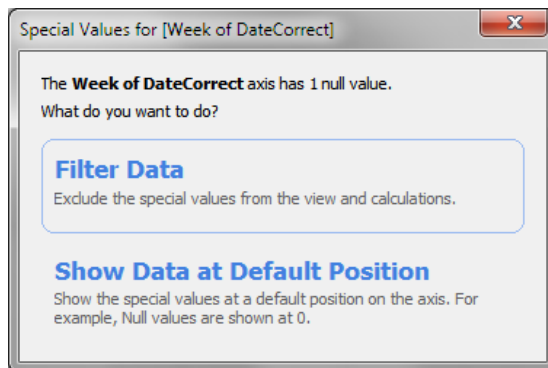
- Now create a similar time plot of the average daily retail price. You can modify the previous chart by clicking the + sign on Week (Date Correct) in the Rows shelf.
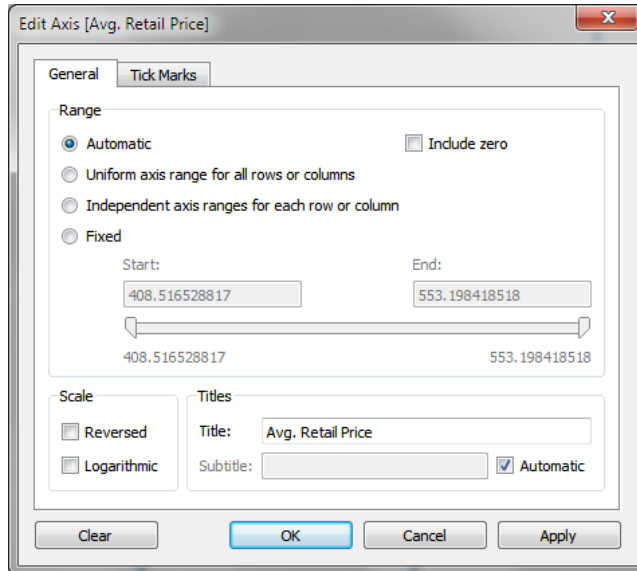
Include in your submission file a screenshot of a dashboard with these two time plots.

Task 2: Including/Excluding the Origin

Note that by default the y-axis origin (i.e. $0.00) is included. This is good practice to make sure we evaluate the fluctuations relative to the magnitude of the series.

To better see the shape of the fluctuations, we can choose to start the y-axis closer to the smallest price in the data.
- Allow the y-axis to exclude the zero origin. Double click the y-axis scale and in the Edit Axis menu that appears uncheck "Include zero".

Include in your submission file a screenshot of a dashboard with these two modified time plots.

Task 3: Temporal Patterns

Does there appear to be a day of week effect in price?
- Break down the weekly average price time plot by day of week (each day should use a different color). Note: because we want to emphasize differences between different days of week, exclude the y-axis origin.
- Hint: add DateCorrect to the color shelf, then click on it and choose More:Weekday.

Try other ways to look at the price changes:
- line chart by day of the week
- boxplots of price per day of the week

Include in your submission file a screenshot of a dashboard with these plots and a sentence commenting on what you learned. Post your dashboard and comment to the discussion board.

Did this also help you identify some missing data? Why is it missing? You can select those data points by marking a rectangle around them, then examine them using Analysis > View Data. In your

dashboard, use Highlight > All Fields. When you choose the Null day of week data in the boxplot, does any data get highlighted in the other plots?

Task 4: Comparing Price Across Stores

Is there a difference between price patterns across different stores?

- Break down the time plot of average retail price by store (e.g. one color per store) There is an interesting pattern to discover here. It is clearer with a daily plot.
- Add a second visualization with a bar chart of total sales per store postcode, and color this visualization also by store postcode.
- Create a dashboard and use brushing between views (Highlight in the Dashboard) to see if all stores have the same pattern (i.e. select a bar of the bar chart to select a store, it will mark the corresponding line in the other display, or select a group of line by drawing a box around a region of the time plot to select the stores whose lines pass by that area.) You can also eliminate the outlier stores with very few sales (select/mark them, then use the right-click menu to filter them out).

Include in your submission file a screenshot of a dashboard with these plots (you can use highlighting) and comment on what you learned.

Part 2: Treemaps (9 points)

You have read about Treemaps in the book but now you can practice creating your own visualizations. To help you we provide a different dataset containing de-identified project management data: project.xlsx

Task 5: Project Owner Spending

- Load projects.xlsx. Choose the categorical dimension "budget type" and the measure "Funds Available" and click Treemap in Show Me. The first level of the "hierarchy" is now by budget type (see Label shelf). The data is now split into 2 rectangles: capital and expenses (i.e. the budget type). The rectangle color and size represent Funds Available.
- Change the chart so that rectangle size reflects Sum(Gross Budget). Change "Budget Type" to "project owner" on the Label shelf. Now each rectangle corresponds to a project owner.

Questions: Can you tell which person manages the largest gross budget? Who is overspending? Include in your submission a screenshot of your treemap and your answer to the questions.

Task 6: Project Owner Spending Per Project

Now add a second level in the hierarchy: add "project name" (hint: add it to the Label shelf). Each rectangle for a project owner is now split to show the individual projects they manage. Some of the people who looked good overall have projects in their portfolios that are running over budget.

**Include in your submission a screenshot of your treemap.**

Task 7: Treemap vs. Hierarchies with Other Charts

- Create a treemap with size by (sum of) Gross Budget and color by (sum of) Funds Available. Use the hierarchy: department, project owner and finally project name.

 Try zooming on part of the hierarchy to see it in more detail. Use filters to see.

While treemaps are powerful to give an overview of the data there are cases where you still want to explore hierarchical data with more traditional visualizations like bar charts:

- Create a three level **hierarchy** of department, project owner and project name [Hint: right-click on Department in the Dimensions shelf > Create Hierarchy; Now drag Project Owner and then Project Name into this Hierarchy). Here is a screenshot of the result in the Data shelf:
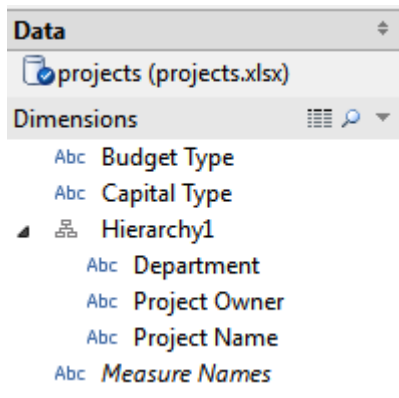
- [Note: Dates are naturally organized hierarchically so you don't have to construct the hierarchy of years, months, days, etc. Tableau takes care of it automatically as long as it recognizes the date format.]
- Next, create a bar chart of (sum of) Gross Budget by the new Hierarchy variable, and color the bars by (sum of) Funds Available.
- Duplicate the bar chart. Now click on the + sign near Department to open the next level of the hierarchy.
- Duplicate the second bar chart. Click on the + sign near Project Owner to open the next level.
- Place the bar charts in a dashboard.

Notice how the bar chart may be easy to compare the size of bars, but you can only see one hierarchy level well. Also, see how much space it takes when we have three levels of hierarchy (requiring scrolling). The treemap is a space filling technique so the larger boxes will always be visible on the

screen even with a large amount of data. Researchers have created treemaps with a million visible items, and users can still easily see the large areas and spot patterns.

Include in your submission a screenshot of your treemap and the dashboard of three bar charts. What do you see from the Treemap?

Part 3: Parallel Coordinate Plot (7 points)

To explore multivariate data you have already used scatter plots with color, shape, size to represent multiple variables. Trellises allow you to see multiple scatterplots or line graphs at once but you can get a different view, using many more variables with parallel coordinate plots.

Tableau 8.1 does not have parallel coordinate plots. However, we can tweak the existing options to create one if we first scale each variable to a 0-100% scale.

Task 8: Multivariate Look at Project Outcomes

Scale each measure to a 0-100% scale (if it isn't already on that scale):
- Click on View Data and examine each of the measure columns. Which of these require scaling?
- For each of the four measures that require scaling, create a new column (Analysis> Create Calculated Field). The following formula will create a new scaled measure by subtracting the MIN and dividing by the range:

100*(AVG([Current Cost])-TOTAL(MIN([Current Cost]))) /

(TOTAL(MAX([Current Cost]))-TOTAL(MIN([Current Cost])))
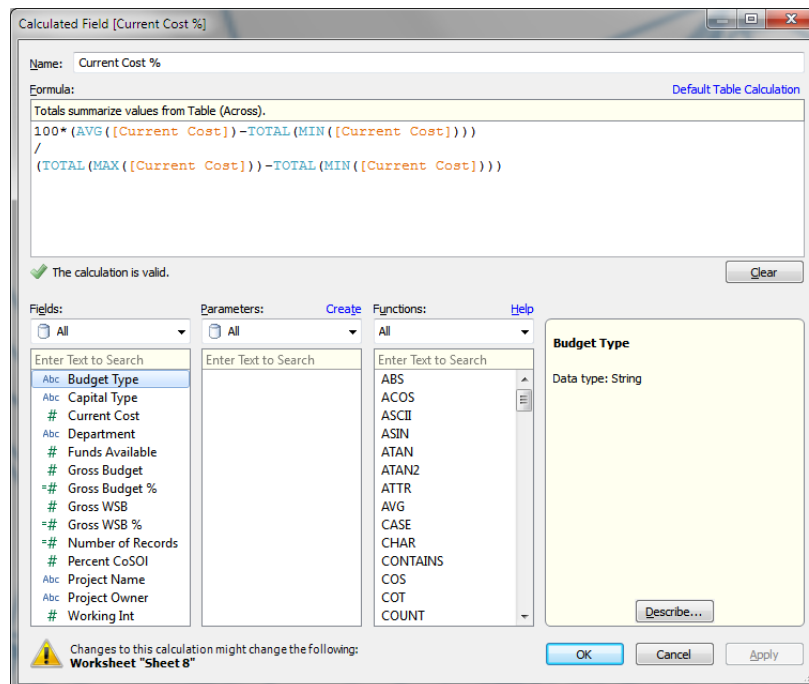
We can now create a parallel coordinate plot of the 5 measures, each of them on a 0-100% scale. Follow the following steps:

1. Start a new worksheet.
2. Drag Measure Names to the Columns shelf. Drag Measure Values to the Rows shelf.
3. In the Measure Values shelf, keep only the scaled Gross WSB, Gross Budget, Funds Available, and Current Cost (in this order). Simply drag the other two out (if you want to keep them, change their aggregation from SUM to AVG)
4. Change the bars to lines by selecting line in the Marks shelf.
5. Lastly, since we want each line to represent a project, drag Project Name into the Detail shelf.

Note: You can drag the right border of the chart's canvas to make the chart wider.

Include a screenshot of your parallel coordinate plot in your submission file with answers to the following questions:
1. Even without really knowing what the data mean or using any interaction **can you spot trends and exceptions?**
2. **Is Gross WSB related to Gross budget?**
3. **How are cost and available funds related?** You can also color the line by something (e.g., department), or highlight some lines.
4. Use marking to select a department and look at all the project measures for that department. **Are some departments standing out?**