Thomas Lederer
Data Analyst Nanodegree

Data Visualization in Tableau Write-Up

Summary

The data presented in the enclosed Tableau story is a set of 113,937 Prosper loans from 2005 to November 2005 – March 2014. The dataset was published and updated on March 11, 2014. Since the vast majority of loans have a term of 36 months/3 years, March 11, 2011 is the last date for which returns can be easily compared to dates prior. Most loans originated after then are "current", which means borrowers are making regular and timely payments to the remaining premium (plus interest) of the loan.

Design: Feedback

In my initial design I was overly concerned with data-ink ratio that many graphs lacked necessary explanatory material. I found that in feedback sessions, the viewer would spend a greater amount of time trying to define the graph's variables than grasping any relation between them. I attempted to address this issue by adding more detailed titles and explicit keys.

Another general revision I made to most slides was to reformat the axes to display variables as percentages rather than fractions wherever applicable. This was a very basic inconsistency throughout the visualization which distracted from the findings I wanted to present.

In other slides, the format or type of graph chosen actually imposed a misleading structure to the data points within it. The dual axis on slide 3 for instance drew the viewers attention to the point at which the lines intersected, suggesting a qualitative change at a specific point in time which should not have been inferred. In the final version I reformatted the slide so the graphs were on separate axes.

Another example was that on slide 7, I originally plotted the default rate on the y-axis (with R.O.I on the x-axis). This gave the appearance that a higher default rate, as higher on the graph, was "better" than a lower default rate. In the final presentation I switched the axes so that ROI lay on the y-axis and default on the x. My intention was that a higher ROI appear intuitively "better" than another point lower on the y-axis.

Lastly, in slide 8, I initially grouped two line graphs on the same dashboard with different time references: loan cycles on one graph, and quarters of loan origination on another. This inconsistency caused a very ambiguous and strained relationship between the two graphs. It is impossible, for instance, to relate the 3rd quarter of 2009 with any specific loan cycle; all loans originated in that quarter have since passed through every cycle. In order to prevent any misleading implications or comparisons, I replaced the line graph with a bar chart.

Below is a full summary of all feedback solicited along with revisions and links to visualizations:

Original Version	**Revised Version**	**Feedback**	**Revisions Made**
Slide 1	Slide 1	Full scale of growth not fully visible	- Include annotations - Set beginning of x-axis to zero
Slide 2	Slide 2	Although cannot really tell without hovering, to what extent loans grew in any state, its clear that growth and decline is shared	No Revisions Made/Not really concerned with any state in particular throughout the analysis - just nature of growth in loans.
Slide 3	Slide 3	Not clear that map gives total values for completed loans, especially since map before allowed users to toggle through years	More explicit titles

		Users eye drawn toward intersection of ROI and default rate which is not intention of graph	Take out dual axes
		Percentages formatted as decimals	Reformatted as percent
Slide 4	Slide 4	Graph made sense to all viewers	No revisions made
		Percentages formatted as decimals	Reformatted as percent
Slide 5	Slide 5	Not clear what "new" and "old" referring to.	Added annotation, added reference line to graph
		Not clear what letters are.	Added more explicit title and legends
		Percentages formatted as decimals	Reformatted as percent
Slide 6	Slide 6	Graph clear in intention	No revisions made
Slide 7	Slide 7	Loans under "old" credit system with higher default rates appear "better" than new loans with lower default rates	Switch axes
		Percentages formatted as decimals	Reformatted as percent
Slide 8	Slide 8	- Not clear where/how loans on top graph y-axis relate to y-axis on bottom graph - "It annoys me that the lines don't join up"	Replace line graph with bar graph
		Percentages formatted as decimals	Reformatted as percent

Design: Graph Selection

Slide No	Graph Type	Reason for Graph Selection
1	Line - Dual Axis	A line graph is the most intuitive representation of time as a continuous variable. I chose a dual axes to emphasize the close relation between the two variables plotted: the total value of loans and the average loan amount. I wanted the viewer to grasp immediately the near-collapse and subsequent rise of the company as determined by these two measures. Since "total loans originated" is the more important of these two, I chose a relatively brighter color for this line.
2	Map - Color	In this graph, I was less concerned with the precise measure of growth in each state than with the encompassing nature of such growth. For this reason, a bar graph - while more precisely readable - would be less approachable when depicting over 50 different categories (/states).
3	Interactive Dashboard: Map - Color; Line - Dual Axis	Since this visualization already depicts two new and less obvious measures (annual ROI and default rate), I choice to summarize these measures over a map the viewer was already familiar with in the previous slide. Since I was contrasting the consistent nature of growth (in slide 2) to inconsistent return (in slide 3), it also made sense to use the same reference point. The graph is interactive to encourage some reflection on the inverse relation between ROI and default rate.
4	Line - Dual Axis; Line	This slide is intended to more discursively and explicitly reinforce some of slide 3. I replaced the map of 50 states with a line graph showing values aggregated more broadly over time only. The reference band is intended to move the viewers up and down the slide and reinforce the relation between the two graphs.
5	Iterative Dashboard: Bar - Color, Line - Color	Line graph used to plot detailed and sudden changes over (continuous) time, bar graph used as more precise visual summary of this data.

6	Bar Graph - Color	Since this slide was intended as supplemental to the slide before it, I chose to simply summarize the results in a bar graph rather than to also plot changes over time. This reduction in detail also allows a clearer color scheme to be used in the bars: distinguishing between the old and new rating system than between each credit grade (as in slide 5) is more fundamental to the purpose of the graph.
7	Interactive Dashboard: Scatter Plot - Color; Bar Graph - Color	I chose a scatter plot to demonstrate the relation/correlation between the two variables plotted (borrower rate and default rate). The addition of color is to demonstrate how this relationship changed under new underwriting standards. The bar graph below was chosen to depict the results of one variable (annual ROI) which I wanted to visualize side-by-side as precisely as possible.
8	Interactive Dashboard: Line Graph - Color, Bar Graph - Color	This is the only slide where time is depicted discretely on a bar graph. I made this choice after a feedback session. Because the line graph above plots time in loan cycles grouped by year, it was misleading to then reintroduce time as a continuous variable below it.