Ikhlas Attarwala
MSAI 339 – Data Science Seminar
10-16-18 – Background Reading

Summarizing "Wrangler: Interactive Visual Specification of Data Transformational Scripts"
by Kandel, Paepcke, Hellerstein and Heer

This paper details the use of a program called "Wrangler". The authors of this paper, from Stanford University and UC Berkeley, discuss the lack of a seamless, data management software. They go on to say how human necessity in programs to inspect and make corrections or changes when integrating data sources can not only be tedious, but near-unending for arbitrary lists and new sets of data to incorporate. Wrangler is a program designed to transform data by a process of "suggesting applicable transform[ations]" based on context, offering "natural language descriptions", and providing an "interactive history". The interface of Wrangler presents many options to visualize and format data differently, but ultimately is designed for great ease of use for all users to work more efficiently.

Firstly, I was very impressed by how the authors took screenshots of their application to guide anyone in basic transforms. The research article essentially became a manual, and something that I found lacking in Stonebraker & Co's paper on the Data Tamer System. The article also does well to go into detail with example scenarios, such as on page 3, when talking about pulling housing crime statistics from the US Bureau of Justice. This is useful for the reader when they're trying to understand the scope of what all is possible when using Wrangler. Finally, to further illustrate how much Wrangler has to offer, the paper goes into depth on a comparison experiment with Microsoft's Excel, one of the world's greatest marketed and most used applications. The comparison set good parameters such as lack of variability between test subjects, similar tasks, but showed how Excel had specific tools for the tasks used to contrast both programs' efficiency. For all tasks, users performed over twice as fast when using Wrangler than Excel. There seemed to be a high preference for visual previews and suggestions.

Something I found lacking in this paper were the limitations discussed. I read 3 topics on how Wrangler did not exceed expectations, and each are related to the user's insufficient experience. This doesn't make for mature writing. In addition, I'm not sure related work appeared in the right place. It could have been placed much later and didn't need as much emphasis as was shown for it. Lastly, I was displeased how this system incorporates many elements from HCI (Human-Computer Interaction), but there's almost no discussion in this paper on the topic.

This paper could use some refinement and further peer-reviewing, but it also doesn't go into great detail on future plans for Wrangler. It momentarily suggests improved tutorials. While they have a great system for extracting and **processing** text, I would recommend looking into Natural Language Understanding to get a stronger feel of recommendations and descriptive analyses.