

Seeing Through the Data: Data Visualization Methods of the Occupational Requirements Survey November 2016

Jenette Muhar and Michelle Myers

U.S. Bureau of Labor Statistics, 2 Massachusetts Ave., NE, Room 4160, Washington, DC 20212

Abstract

The Occupational Requirements Survey (ORS) is the newest survey conducted by the Bureau of Labor Statistics (BLS). Created as a way for the Social Security Administration (SSA) to be able to determine what abilities workers need to have in order to perform a job, we rely on workers, managers or HR specialists, during collection, to provide us with levels and/or durations of time spent doing things like lifting, driving, making decisions, etc. The ORS, although similar in some ways to the DOT, measures a unique mix of elements. This led to the need for us to provide data reviewers and the validation team with some context to better understand the relationship between elements. In response to this need, we have enhanced our review systems by integrating data visualizations to aid reviewers in identifying outlier data and pinpointing relationships among variables that can be brought back to the field, increasing the efficiency of the process. This paper details the usefulness of data visualizations, the process of creating relevant visualizations, and the outcomes we have seen since the implementation of the data visuals in ORS micro data review and estimate validation processes.

Key Words: Review, Validation, Data Visualization, Graphs, Dashboard

1. Introduction

Over the past four years the Bureau of Labor Statistics has been creating a new survey to be used by the Social Security Administration (SSA). The survey is called the Occupational Requirements Survey (ORS). The ORS is designed to collect data on the types of physical and cognitive activities that are required for workers across the nation [1]. In general, ORS will provide a wealth of information about job characteristics and will serve as a resource of occupational requirements for the general public and as a source in the SSA adjudication process. Currently the Dictionary of Occupational Titles (DOT) [2] is the most comparable data to what is being collected in ORS, but it has not been updated since 1991. Because of this, much of the data is not directly comparable and the ORS data review team has been posed with the task of finding objective ways to accurately review the collected ORS data. For this, we have found part of the solution is the use of data visualizations.

With new easy-to-use data visualization software, analysts are now able to create and use their own data visuals without having to go through developers. This is important because not only do the improved software systems make the analysis more efficient but with the exponentially increasing amounts of data, data visualizations allow information to easily be reviewed, compared, and shared even with others who may not understand how it was prepared. Data visualizations have been extremely useful to those working on ORS, particularly to the micro data review and validation teams.

In October of 2015, the BLS Office of Compensation and Working Conditions created a team whose goal was to put visually-displayed data in the hands of reviewers quickly and without requiring reviewers to pull data from the database. Our team achieved this by building a series of dashboards that would serve to provide context for reviewers when analyzing collected data. The results of piloting the data visualization dashboards have been promising. They have given reviewers context in which to find flaws in the collected data. Additionally, the dashboards have simplified some of the review processes and have improved the efficiency of estimate review by providing a one-stop-shop for estimates and their corresponding micro data. The visualization methods chosen and their usefulness in the ORS process will be detailed in this paper.

2. Project Background

The ORS visual dashboard serves to aid users in the two major sections of the review process: micro data review and estimate validation.

2.1 Micro data review

Once the data are collected from the field and captured in our collection system, each collected establishment is run through a program we refer to as the roulette wheel. This roulette wheel assigns an establishment's data to a specific type of review and subsequently to a specific reviewer, to be evaluated based on edits unique to that particular review type. While reviewing, the reviewers have the option to question the data coded, ask for additional documentation for a coding, or verify that the data is correct as coded. Some of the types of review done are called Secondary, Targeted, Technical Re-interview Program (TRP), and Staff Development Analysis (SDA). These review types will be briefly described below. A more in-depth look at the different types of review done in the ORS survey can be found in 'Occupational Requirements Survey (ORS) Data Review Process' as well as 'Building Quality Assurance for the Occupational Requirements Survey'. [3][4]

- Despite its name, secondary review is the main type of review in which approximately 60% of establishments are assigned. In secondary review, edits are triggered that look at the relationships among the data elements (i.e., it is coded the worker climbs ladders but also that time spent sitting is 100% of the work day.) These edits are based on a combination of patterns we saw during feasibility/pre-production testing, data that comes from O*NET [5], and professional expectations. In general, they are broad edits applied across all jobs with only a few select edits that apply to particular occupations. Along with the edits described which check the relationships of the data elements, we also have "certainty edits" which are edits that trigger for every observation. In secondary review the certainty edits that trigger have reviewers verify the SOC for each observation. The secondary edits described above are also applied to the targeted, TRP, and SDA review which will be discussed next.
- Targeted review is assigned to approximately 20% of collected establishments. When an establishment is assigned to targeted review, in addition to reviewing the secondary edits that are triggered, the reviewer also looks at specific variables. The specific variables they look at differ for each establishment. Variables to be reviewed are randomly assigned when a sampled establishment is run through the roulette wheel. There are specific clusters that may be assigned and each cluster is comprised of five variables that need to be reviewed. For targeted review, each job that is selected will be assigned two clusters so a total of ten variables will be checked per job selected. The reviewer will look at each of those variables to make sure the coding appears reasonable based on what is intuitively expected and the documentation provided.

Essentially, targeted review is spot checking to make sure the data are coded appropriately. In addition to the targeted edit clusters, there are also certainty edits particularly for targeted review. These certainty edits have reviewers check the duration of time sitting vs standing and SVP level for each of the observations.

- The third type of review is the Technical Re-interview Program (TRP) in which the reviewer re-contacts the establishments surveyed to verify data. The purpose is to review the quality and integrity of the collected data. During the re-contact, reviewers ask the respondents questions on a random selection of occupations and data elements that were asked during the initial survey to ensure consistency of the data. TRP is assigned to approximately 5% of collected data.
- The remaining roughly 15% of establishments are assigned to Staff Development Analysis (SDA) review and those reviewers look at all of the approximately 70 collected elements individually from an establishment. The goal of SDA is to analyze the collected data and confirm that the data coded for the variables does not contradict what is coded for other variables, assist in staff development and support, and ensure survey procedures are followed. They are essentially looking at everything that has been collected for a specific establishment as part of general review.

All of the micro data review discussed previously is focused on the data within one collection establishment. The other kind of review that is done looks at data across all collected establishments to find data points that are outliers and should have their coding reviewed. Since the ORS is a new survey, we have employed two complementary methods to help reviewers determine which data points are true outliers. After going through one of the four types of micro data review, all the micro data is again compiled and run through a SAS outlier review program. Ideally we would have the data weighted prior to being run through outlier review, but as it is still early in the ORS cycle this process is currently being conducted sans weights. Once a method has been developed to determine how to incorporate weights for the data, the program will be updated and incorporated in the system.

First the data are run through a program that compares the data point of an element to other data points for the same occupation (SOC) [6]. Within that SOC, the program flags the numeric data points that fall above the 95th percentile or below the 5th percentile of a given element. For the data that are collected by categories, the program flags any category as an outlier that makes up less than 5% of the collected data, sans any unknown, for that element. The parameters for this program are still being reviewed and other potential ways of statistically determining an outlier are being considered.

The above program flags quite a few data points as outliers. So in order to narrow the scope of outlier review and take other factors into consideration, such as distribution of the data, establishment size, or industry; we have created a visualization dashboard. This dashboard shows the data that are being flagged in the program against all other collected data points. By using this, reviewers are able to see where the flagged outliers fall in relation to the overall distribution of the data. They can also filter by factors such as establishment size or industry (Perhaps in-home nurses have different requirements than nurses at a hospital). They can compare variables to one another (Perhaps a laborer lifts less often than normal but they are also required to be on a ladder longer). All of these factors are important but vary in importance on a situational basis and therefore cannot necessarily be determined solely by a statistical program.

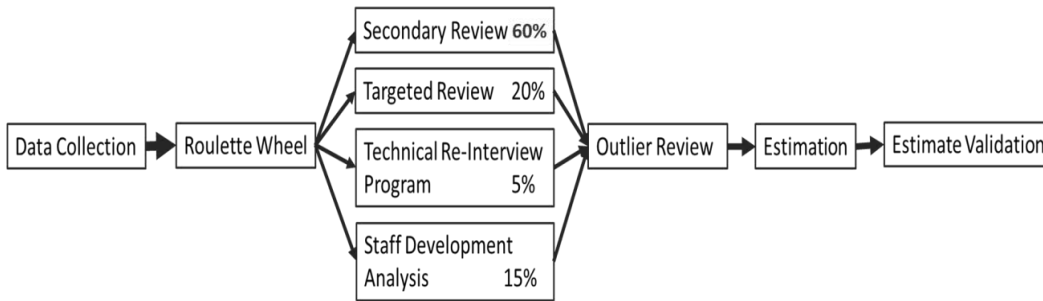


Figure 1. Outline of the process from initial collection through the estimates being validated.

2.2 Estimate Validation [7]

After the micro data has gone through the process of being reviewed, and possibly questioned and changed, it is then run through estimation. In the case of ORS, we will have about 500,000 different estimates. It is then up to the validation team to review the estimates and validate that the estimates we are providing make sense. Typically this could be done using a spreadsheet or having validators read through the different estimates manually but with so many estimates it would be quite time consuming and inefficient to perform this type of review. Due to the large number of estimates that require validation, we again turned to data visualization to help with the process.

We have created a validation dashboard with multiple views which enables the validators to find answers to all the questions they have about the estimates in one place. The validators can see and compare the value of the estimates within a particular SOC, find estimates that are being driven by imputed data, and view the changes in weights, imputation rates, and suppressed data from one cycle to the next, all at the same time.

Through the use of visualizations we have also made it possible for the validator to go from viewing the estimates, to viewing the underlying data that is feeding each estimate, and then viewing whether particular data points of interest have been questioned during micro data review. Having one place that validators can go to and have all the information they need in order to quickly see anomalies in the estimates should make ORS validation more efficient than it would be otherwise.

Although the dashboards serve different purposes, they work together as one for a common goal. Below we will explain the thought process behind our choices in development and what factors led to a cohesive visual process.

3. Our Approach

There were four overarching principles that we tried to maintain throughout the dashboards.

3.1 Consistent

First we focused on consistency; although the majority of those using the dashboards are micro data reviewers and would only need to use sections of the dashboard, the validators would be required to use all of them. For this reason we made sure that all of the dashboards were consistent in their style and that the flow from one page to the next was as seamless as possible. Each of the dashboards we created were set up in the same way, the filters, graphs and functions were placed in the same locations on the page and the clicks function

in the same manner on each section of the dashboard. This is helpful for the user but also beneficial for the maintenance of the dashboard. For the user it shortens the learning curve by making it so that no matter which dashboard the user is on they know how it can be used. As for maintenance, if something is found to be wrong or needs updating the changes for one portion will be the same changes needed for the other portions as well.

3.2 Dynamic

Secondly, we wanted to make sure that all the dashboards were dynamic. By dynamic we mean not only do users have some options for changing the particular graph they are looking at but also that live data was feeding the graphs so they will adjust in time with the data being collected.

Whether visualizations are static or dynamic play a large role in their usefulness. A static graph is one that does not change, it is essentially created by one person and other people can view it but the data it is showing does not change. Dynamic visualizations are ones in which the user is actively engaged. Dynamic visualizations have grown substantially in their use for disseminating data to mass audiences because they peak users interest and make them feel as if they are part of the data. It is also useful for Exploratory Data Analysis where analysts are exploring the data to find possible relationships and things to test in the future. In our case, we have used dynamic data visuals to conduct Exploratory Data Analysis. The dashboards we have created have the chart types fixed, but the user has the ability to select the level of data they want to explore. This is important because the questions each user has about the data may be different and by having dynamic charts these various questions can be answered.

We have provided users with drop-down selection menus where they can choose the exact data they want to see. For the micro data review dashboards, users can chose to look at a specific job, establishment, industry or any of the various worker characteristics. They also have the option of simply viewing an aggregate of all the data. In outlier review, this ability makes a huge difference. If something is triggered as an outlier for a specific SOC, the reviewer can then go into the data and answer questions such as, is this something that is not often seen in the occupation but is normal for this industry? Or maybe a cluster of outliers whose relationship between two variables stand out compared to all the aggregate data, but when compared to businesses with less than 50 employees the data looks completely normal.

Another aspect of the dashboards being dynamic is that we connected the dashboards to live data. By connecting them to live data, we enabled reviewers to see their exact data point against other data points at any given time. This is vital since data are constantly being collected from the field and can be changed at any point during the review process. The inclusion of live data allows users to see up-to-date information any time they use the dashboards.

3.3 Understandable

Third, the dashboards needed to be understandable. We built the dashboards with the intention that if a new employee was brought on they would still be able to use and understand the dashboards even with no real knowledge of the data that was being displayed. All of the dashboards that were created during this project were designed to function as a web page. The general set up and clicking functions are more useful this way because by this point all new hires have used a web page at some point. Providing users

with a tool that they feel familiar with right away cuts down on the time spent training and shortens the learning curve. Both of these points are essential in creating a truly useful tool.

3.4 Concise

Lastly, but most importantly, we made sure that what was shown was what users would absolutely need to see in order for the tool to be useful. Making sure we can see all the pertinent information and not to include more information than the user required was an important feature of the dashboards. Cutting out extraneous information narrows the scope of the data to help users focus in on what they should be looking at.

Keeping these four things in mind we designed and developed a series of dashboards to help users with situations described above. The project has initially been done using Tableau Desktop [8].

4. The Visuals

For each of the visualizations we created in the dashboards, we let the data and questions we were trying to answer drive what the dashboard would look like. Certain types of graphs answer particular questions better than others. Below we will explain the types of graphs we chose to use, the reasons we chose them and the questions that they are answering in the ORS review process.

We started by determining what questions reviewers would be asking while they were evaluating the data. We have chosen to explain the types of visuals that answer each question and how their application can help each of the stated processes.

4.1 Question: What are the values of the data/estimates?

In the ORS survey there are two ways we collect variables: some are collected as a numeric variable (the majority of which are measured as a duration of the work day) and others are selected from pre-set categories. We will be referring to these types of variables as duration variables and categorical variables, respectively, throughout this paper. In addition to the micro data, we also review estimates, which are calculated as numeric variables. Duration variables and estimates are continuous variables while categorical variables and edit data are discrete. This distinction between the types of variables has led us to use two different types of graphs to visually answer the same question.

To visualize how the continuous variables are being coded, we chose to use a box-whisker plot because it helps reviewers see the distribution as well as any potential outliers. A box-whisker plot generally shows data points in a straight line along with a box which shows the 25th to 75th quartiles over the data points. This box is divided in color by the median. The whiskers on our graph are set to 1.5 Inter Quartile Range (IQR) which is a standard setting in Tableau. The IQR is the difference between the 25th and 75th quartile so having the whiskers go 1.5 the length is a fairly accurate representation of where the data are expected to fall. Being able to see not only the box and whiskers but also each individual data point is very useful to encourage those who may be less trusting of relying on data visualizations and would not embrace seeing just the aggregated distribution. In order for the user to be able to see each individual data point, we had to jitter the data because without that multiple points may have the same value and would appear as one point on the graph. We used color to show what the SAS program flagged as an outlier because it gave us an easy comparison for whether the SAS program coding was flagging what we would visually consider an outlier. We included on the tooltip [9] other necessary information

such as worker characteristics (Establishment ID, Percent coded, etc.) that would be of use..

For elements that are measured as a categorical variable (i.e., Do they drive?; yes, passenger; yes, other; no), the box whisker plot is not an ideal way to view the distribution. Another option was to use a stacked bar chart, which is essentially a vertical pie graph. Although this method seemed most useful, it is important to keep in mind the audience who will be using the dashboard and in particular what they will be using it for. In our case the main purpose of seeing the distribution of the data was to find anything that would be a potential outlier. The users needed it to be very clear where they may need to focus their attention. With the stacked bar chart, any potential outlier would be so small it would be easily missed. Instead, we opted to graph categorical variables on a standard vertical bar chart where each category was its own separate bar. We applied the same concepts as we had with the box-whisker plots, highlighting those categories which made up less than 5% of the total (to show potential outliers) and included all needed values on the tooltip. The bar chart again, like the box-whisker plot, helps us to see the general distribution and potentially even outlier data.

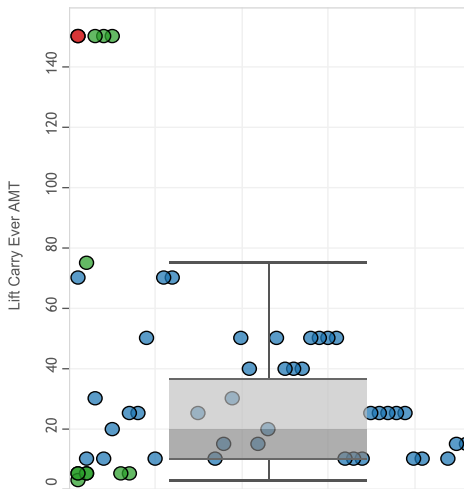


Figure 2. Box-Whisker plot showing numeric data distribution. The red and green dots indicate data that was flagged as an outlier in the SAS outlier program.

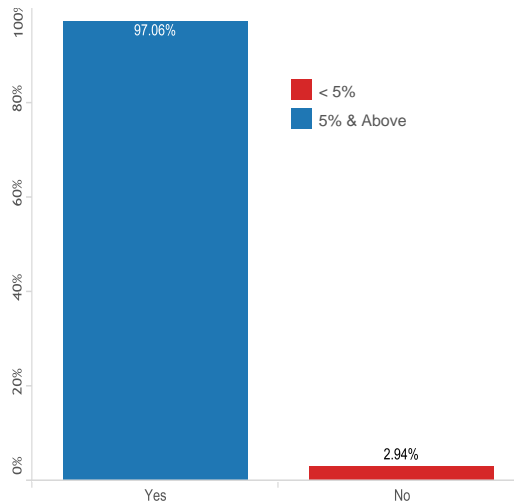


Figure 3. Vertical bar chart showing data element collected as category where the red indicates potential outlier data.

By using these types of graphs, micro data reviewers are helped in a few ways. First, they are able to see the distributions of the data that has been collected so far giving them context for determining whether the data point they are reviewing makes sense or not. For example, in figure 2 the reviewer would clearly be able to see that for this particular occupation and worker characteristics any time a worker is coded as lifting over 80lbs is highly irregular. Likewise in figure 3, if they were reviewing this variable they would be able to see that for the given data element it would be highly unlikely the respondent would have answered 'no'.

Second, by using either the box-whisker plot or vertical bar graphs micro data reviewers are also able to clearly see potential outliers within an occupation. This is useful in any

type of review but specifically in outlier review. By viewing the un-weighted data like this you can see which points stand out as being substantially different than the rest. This will not catch all the outliers, but it will help those reviewing micro data catch extreme values before such data are included in estimation. Using this review, we have a chance to question and find out what is going on before changes are no longer possible and micro data or estimates have to be suppressed [10].

4.2 Question: What are the relationships between the data/estimates?

“No man is an island” and neither is a variable. Most of the ORS variables have some sort of relationship with another. For some it may be that they cannot occur simultaneously (i.e., sitting and standing.) While for others the relationship is that if one is coded the other must also be coded (i.e., if driving is coded vision must also be present) and there are others where we do not yet know the relationship. Knowing these relationships early on in the ORS process helps reviewers to provide better data in the long run. When validating estimates, being able to see the relationships has a different purpose. It is important to see the relationship between the weight and imputation of data in an estimate to determine whether certain data should be excluded from an estimate. If a particular data point has a high weight, it is potentially driving the estimate. If that data point has been imputed, we may not want to include it in the final estimate.

In order to see the relationship among multiple element values in the micro data and between imputation rates and weights of the estimates for continuous variables we used a scatter plot (see figure 4). In general, a scatter plot is a good way to see relationships and can be produced in almost any program. In this case it was possible to use other types of graphs that would have been more exciting or visually appealing. However, it was more important that the graphs be clearly understandable. In order to show the same relationships for the categorical data, a scatter plot was not a viable option since all the points would fall on the same value. Instead a heat map was used (figure 5). A heat map is a grid where each square represents a combination of variable X with variable Y. It is color coded, where the color is indicative of the number of data points in the given combination. So the darker areas on the graph are more concentrated with data; thus, the combination is more frequently collected. In figure 5, for example, there have been no observations that have been coded “No” for the element on the Y- axis that have also been coded either “A” or “B” on the X- axis element. Similarly, the most frequent combination of the 2 elements is when the Y- axis element is coded as “Yes” and the X- axis is coded “C”.

Both the scatter plot and heat map excel at showing patterns in the data. With the scatterplot the user is able to see how the data clusters based on where on the graph the most points are falling. For the heat map they are able to see those same clusters through the color of the squares.

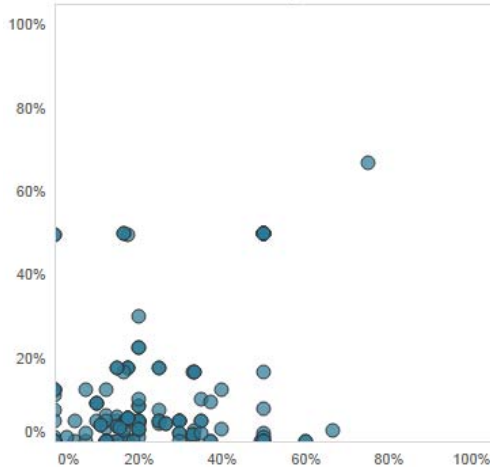


Figure 4. Scatter plot of two variables measured in duration of the day.

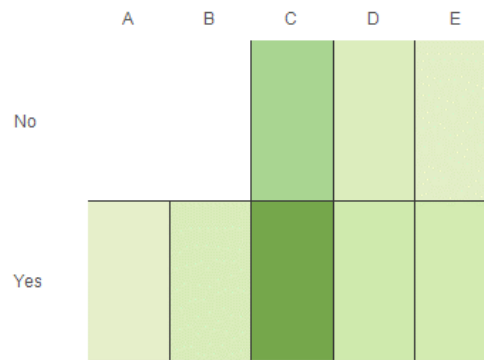


Figure 5. Heat map of two categorical variables where the darker green is indicative of more data being coded with that combination.

4.3 Question: What has changed?

With ORS data it seems likely that there will be relatively little change from one cycle to the next due to the nature of the data being collected. Therefore, any drastic change will be very important to take a close look at.

For micro data review, the biggest reason there may be changes across time is that the definitions of some of the variables are still being refined. It is important that the reviewer is not comparing data from the current time to data from a time the definition was different. For this we have included a drop-down box where the user can select a specific collection cycle or choose to view all the data that has been collected so far.

For estimate validation we are interested in seeing the changes in the imputation rates, the weight of estimates, and the number of suppressions over time and across multiple collection cycles. In order to easily display this, we used a vertical bar chart to show the change in the percent value of the estimate that has occurred since the previous cycle. We visualized this by using the concept of a bar chart but replacing the bars with circles to show the specific points. This way we were able to see all of the estimate values without any overlap. We made a drop-down filter that controlled what SOC is being viewed and on the X-axis we put both the collection cycle as well as the estimate characteristics. In figure 6, the collection cycle number is shown on the bottom X-axis and is either 701 or 702 while the characteristics of the estimates, or sub-cell IDs are shown on the top of the graph. The sub-cell IDs are particular characteristics of a SOC and show breakouts such as whether the estimate is for full-time/part-time workers, or union/nonunion workers. That way we could easily see how a particular estimate changed from cycle to cycle. We also used color to differentiate the type of estimate which made things clearer for the user. The user is also able to see the changes in imputation and weights by clicking on any of the point estimates on the graph. Once they interact by clicking on a point, two additional graphs display below showing the weights and which data points were collected or imputed for the data contributing to that estimate for each cycle that it is available. Having this visual dashboard of the estimates will help understand and explain the estimates, particularly early on in this survey.

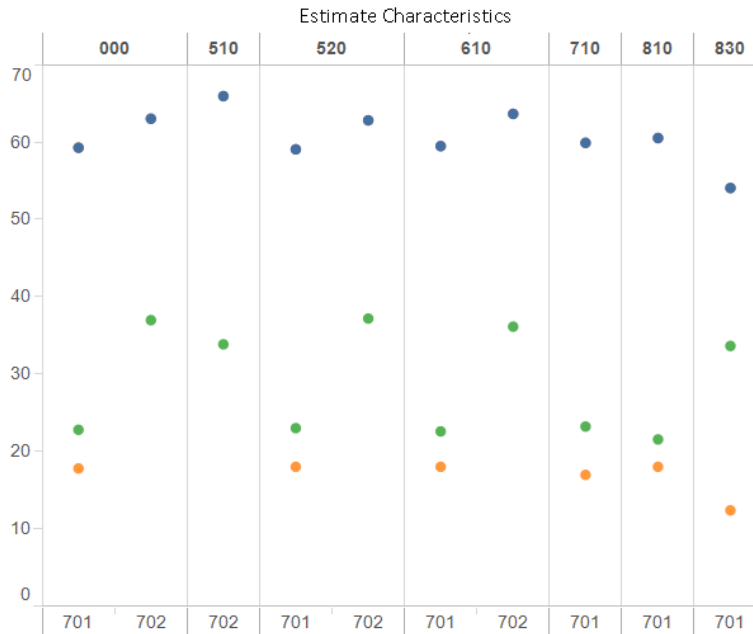


Figure 6. A vertical bar chart where the bars have been replaced with circles and the color indicates the particular estimate.

5. Conclusion

As with all things in ORS, the visual dashboards are new and consistently being re-evaluated to fit the needs of the review and validation staff. So far the benefits during review are promising. The addition of the data visualization dashboards has led to ORS micro data reviewers having more confidence in the data points they choose to question and spending less time looking at outliers. The estimate validators have saved time and energy by not having to flip between multiple spreadsheets of estimates and micro data. And we will soon have a place for managers or quality reviewers to track the effectiveness of our edits system. The use of the data visualization dashboards designed for micro data review and estimate validation will help increase the efficiency of the ORS production processes and will provide opportunities for data exploration going forward.

In building a series of dashboards that focus on users' needs, are intuitive, consistent and dynamic; we have been able to add context to data that essentially has no other comparison. In the future we plan to continue to use the visual prototypes and to further develop visual systems to aid in review activities. By adding these visualization tools into the analysts' repertoire so early in the ORS life cycle, we believe we will be able to produce increasingly more accurate estimates in more efficient ways, adding to the success of the Occupational Requirements Survey.

Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics or the Social Security Administration.

References

- [1] More detailed explanations about what the survey entails as well as its purpose can be found on the ORS website <http://www.bls.gov/ors/>
- [2] U.S. Department of Labor. (1991). Dictionary of Occupational Titles (Rev. 4th Ed.). Washington DC: U.S. Government Printing Office. <http://www.occupationalinfo.org/>
- [3] Meharena, Ruth. 2015. Occupational Requirements Survey (ORS) Data Review Process in *JSM Proceedings*, Government Statistics Section. Alexandria, VA: American Statistical Association.
- [4] Brown, Karen and Harney, Tamara. 2015. Building Quality Assurance for the Occupational Requirements Survey in *JSM Proceedings*, Quality and Productivity Section. Alexandria, VA: American Statistical Association.
- [5] U.S. Department of Labor, O*Net Online, <http://www.onetonline.org/>
- [6] Occupations are categorized using the Standard Occupational Classification (SOC) which is used to classify occupations based on the tasks they perform as opposed to title matching. For ORS occupations are classified using an 8-digit code which is composed of the 6-digit SOC and 2-digit suffix defined by O*NET. See Standard Occupational Classification website, <http://www.bls.gov/soc/>
- [7] Smyth, Kristin. 2015. Validation of Estimates in the Occupational Requirements Survey: Analysis of Approaches in *JSM Proceedings*, Government Statistics Section. Alexandria, VA: American Statistical Association.
- [8] Professional Edition Tableau Desktop version 9.0 Patent- <http://tableau.com/ip> ©2015 Tableau Software, Inc. and its licensors. All rights reserved.
- [9] A tooltip is a message that appears when a cursor is positioned over an icon, image, hyperlink, or other elements in a graphical user interface. - Oxford Dictionary
- [10] Due to the youth of this survey a hard suppression process has not been established but it is fair to suspect that suppression for the ORS will be similar to the process of the National Compensation Survey.