ABSTRACT

Walt Disney once said, "Of all of our inventions for mass communication, pictures still speak the most universally understood language." Using data visualization to tell our stories makes analytics accessible to a wider audience than we can reach through words and numbers alone. Through SAS® Visual Analytics, we can provide insight to a wide variety of audiences, each of whom see the data through a unique lens. Determining the best data and visualizations to provide for users takes concentrated effort and thoughtful planning. This session discusses how Western Kentucky University (WKU) uses SAS Visual Analytics to provide a wide variety of users across campus with the information they need to visually identify trends, even some they never expected to see, and to answer questions they might not have thought to ask.

INTRODUCTION

One of the best things about data visualization today is how dynamic it is. Platforms like SAS® Visual Analytics allow one visualization to meet the needs of multiple audiences with just a few mouse clicks or taps on a mobile device. If a visualization is set up correctly, it can tell a unique story to each individual user based on the user’s interests and needs.

This paper will take you through the thought process we used at WKU to develop our data visualization system in SAS Visual Analytics. At the same time, we will provide examples of our visualizations currently used by users across campus to make decisions. While any reader interested in using Visual Analytics will benefit from this paper, those with some experience using Visual Analytics Explorer and Designer will benefit the most. Additionally, while readers in postsecondary education will be more familiar with the examples used in the paper, the data visualization processes and Visual Analytics functionality apply across industries.

A VARIETY OF USERS AND NEEDS

Enterprise data visualization systems have to address the needs of many different audiences. Understanding the data needs of each audience is crucial to the success of the implementation. At WKU, we have a variety of users at different levels of the organization. At one level, we have our President and all of the Vice Presidents/Chief Officers. This executive group wants to see data at a high level, what we often call the “40,000-foot view.” They have a limited number of indicators they follow, but they want to see them for all students, faculty, staff, or departments at the university. They like to see trends, but they also need to be able to see what is behind those high-level trends, particularly if they are not going in the desired direction.

The next group of users includes the Provost and Vice President for Academic Affairs and the Academic Affairs Executive Team. This group still likes to see data at a high level, but they have a large number of indicators they want to track and they typically focus on students and faculty—the academic side of the University.

The last group includes everyone else—Academic deans and department heads, advisors, program coordinators, and university directors. This group wants both high-level and detailed data, to track a moderate number of indicators, but their focus is highly specific. For example, staff in the Office of Admissions focus on the application process for incoming freshmen and transfer students; faculty and staff in a specific college or department focus on their majors from application to graduation, while staff in Residence Life focus on students living on campus.

FLEXIBILITY FOR USERS

The first way we tried to meet the needs of our users was by providing as much flexibility for them as possible. We produced our visualizations at the university level, but provided numerous filters for each
visualization. This allowed users to access trends at the highest level, but also see trends for their specific populations and discover which groups were contributing to changes in the trends.

Because visual real estate is at a premium, we chose Prompt Containers (See Display 1) to hold numerous List Controls (See Display 2). By using prompt containers, users can open up the container, make their selections, close the container, then see the results on the full screen. While Button Bar and Drop-Down List Controls are more visually appealing, they allow users to choose only one value at a time. We chose List Controls to allow users to select multiple values for each filter, which provides them with maximum flexibility.

Display 1: Prompt Container

<table>
<thead>
<tr>
<th>Level</th>
<th>Division</th>
<th>Enrollment Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>UG</td>
<td>Lower Division</td>
<td>Part Time</td>
</tr>
<tr>
<td>GR</td>
<td>Upper Division</td>
<td>Full Time</td>
</tr>
<tr>
<td>Other Undergraduates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduates</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Display 2: List Control Examples

When we had natural hierarchies in our controls, we created interactions between the List Controls to cut down on the need for users to scroll through irrelevant values. For example, when we used filters for college, department, and major, we created an interaction between college and department and between department and major (See Display 3). This allowed users to choose their college, which filtered the departments to show only the departments within that college. Once users chose a department, it would filter the majors to show only the majors in that department. While you can use a Report Prompt to filter a Section Prompt, which automatically filters a prompt located in the canvas portion of a report tab, it is not possible to create these interactions between controls if the controls are all located within the Report Prompt area or within the Section Prompt area of the report. It made more sense for us to put all of the List Prompts in a Prompt Container located in the canvas portion of the tab, and then create the interactions between prompts and other objects on the canvas.

One of the biggest challenges in trying to report on postsecondary students is whether to report the number of students, which we call "headcount" or the number of majors, of which each student can pursue up to four and each of the departments and colleges for those majors want to see data for the student. Visual Analytics allows you to do both and let your user choose which way they want to see the data. To do this, we set up our data to have one record for each student for each of his or her majors. Then, for each major, we created a variable called MAJCOUNT that showed which major it was. We chose to label our first major "Headcount (First Major)", then label the second major "Major 2," the third major "Major 3", etc. Next, we added a Drop-Down Control to our report and placed MAJCOUNT into the control. We labelled the control "Select for Headcount/Clear Filter for Majors." Most importantly, we placed a filter on the control that showed only "Headcount (First Major)." This allowed users to see only the "Clear Filter" and "Headcount (First Major)" options (See Display 4). By adding this control, users select "Headcount (First Major)" to show head count, or "Clear Filter" to show all majors. We then trained our users to understand that showing all majors is appropriate only when viewing reports at the major level.
DIFFERENT REPORTS FOR DIFFERENT USERS

Another way we met the needs of our users was to provide different reports based on the level of data and number of indicators the users needed to see. For our President and Vice Presidents/Chief Officers, we produced one high-level report that focused on the specific indicators they wanted to track. This report updates automatically every weekend with the current week’s data, which allows the executives to discuss any trends at their regularly scheduled Monday morning meeting.

The majority of our reports are available to all users and update automatically each night or each semester, depending on the type of data and the frequency with which it changes. These reports focus on topics such as student enrollment, degrees awarded, student persistence, student success, and recruitment. Some users focus on the overall trends, while others want to see the detailed data beneath the trends. We provided reports that allow users to see both.

The final way we met users’ needs with different reports was by developing a very limited number of specific reports for specific user groups. For example, the Academic Affairs Executive Team wanted to explore the factors influencing the number of credit hours earned at the time of graduation. Unlike a static report, Visual Analytics allowed them to explore the many dimensions of the issue and find patterns they did not expect to see. Note that these specialized reports are the exception, rather than the rule. Our goal was to develop as few reports as possible to meet the needs of all of our constituents. We thoroughly evaluate whether or not to develop specialized reports based on the number of users that will benefit from the report, whether the report is in line with our mission, and whether Visual Analytics is the best tool to answer the need. The last thing we want to do is to have to maintain numerous reports that are rarely used.
STORYTELLING WITH DATA VISUALIZATION

ENROLLMENT

Once you have fully taken stock of your users’ needs, it is time to start developing reports that tell a story. One of the major stories at our institution revolves around enrollment. As a public institution that receives less than 17% of its budget from the state, we rely heavily on tuition paid by students. From 1997 to 2012 enrollment grew 45% from 14,543 to 21,124. We were used to enrollment growth every year. Then, for each of the next three years, we saw a decline in our enrollment, corresponding to a decline in revenues. Information about enrollment and the factors driving the decline were in high demand. One of the charts we developed to track enrollment during each week of the enrollment cycle is shown in Display 5.

Display 5: Fall Enrollment by Week 2012-2016

For the line chart shown in Display 5, we numbered the weeks in the enrollment cycle so that Week 0 (Zero) represented the week classes began in the semester, Week 9 represented the census, the week we freeze our data files for reporting purposes, and Week -9 represented nine weeks prior to when classes began. By standardizing the weeks in this manner, we could lay the years on top of each other to see if we were up or down during the current semester, compared to the same time in the enrollment cycles for the past four fall semesters. However, by standardizing the timeframe, we lost the exact dates for each week. To allow users to see the dates associated with each data point, we added three numeric variables to the data: Month Extracted, Day Extracted, and Year Extracted. We then added those variables as measures in the Data Tip Values for the chart. When users hover their mouse over a data point, as shown in Display 5, they can see the date associated with that data point, which in this example is April 30, 2016.

While this chart is great to track overall enrollment patterns, it is particularly useful in showing sudden spikes or drops in enrollment for selected populations. While they can show actual changes over time, these sudden changes can also be indicative of other things. First, they can show when you have missing data. Capturing and formatting large amounts of data for every week in an enrollment cycle is no easy task. When values suddenly drop to zero for a week or two, and then shoot back up to normal levels, you can easily see which weeks are missing data.

Sudden changes in longitudinal data can also be an artifact of a process. One Monday morning we received a call from one of the Vice Presidents asking about a sudden increase of 500 part-time students. Because he tracked enrollment data closely, he was concerned about this sudden increase. Using graphs like those shown in Display 6, we were able to explain what was really going on. Before each of
our freshmen orientation sessions at WKU, staff in our Academic Advising and Retention Center (AARC) pre-register incoming freshmen for six to eleven hours, which enrolls them as part-time students. On the day of orientation, students meet with their advisors and finish registering for the rest of their classes, usually bringing them up to full-time status (12 or more hours). As you can see from Display 6, there are peaks in the number of part-time students throughout the enrollment cycle, followed by drops in the number of part-time students that correspond to increases in full-time students. Our vice president happened to look at enrollment for part-time students right after AARC had enrolled students, but before orientation had occurred.

![Display 6: Fall Enrollment by Week and Enrollment Status 2012-2016](image)

Sudden changes can also be just that, sudden changes. This past fall, we noticed a sudden increase in students taking all of their classes online, shown in Display 7. While this chart is good for showing relative changes, we used another visualization (See Display 8) to investigate what was causing this sudden uptick in students taking all of their courses online. The table in Display 8 takes the current week in the semester, which, at the time, was two weeks after classes started, and compares data for the same week for each of the past four fall semesters. It provides a spark line, a traffic-lighted 1-year difference and a traffic-lighted four-year difference. Besides being able to filter on numerous variables, the report allows users to dynamically substitute categories in the chart to examine different variables. For a more thorough explanation of how to do this, Lui (2015) provides an easy-to-understand explanation of how to dynamically substitute categories and measures into graphs and tables in SAS Visual Analytics.

To investigate the dramatic increase, we filtered the table to show only those students who were taking all of their courses online. Then, we started selecting categories to see if we could find a group of students that stood out. What we found was a large increase in high school students taking courses online. This large increase made sense for two reasons. First, high school students typically take one, sometimes two courses a semester, so 100% of their courses typically means one or two courses. Additionally, the
Governor of Kentucky had recently instituted a scholarship fund so that high school seniors could take two college courses at no cost to the student. As suggested by the chart, many took him up on his offer.

Display 7: Enrollment by Week for Students Taking All Courses Online

Display 8: Current Week Enrollment Compared to Previous Similar Weeks by Classification

PERSISTENCE

Another key story at many postsecondary institutions is student persistence. It is not enough to enroll students—we also want to know if our students are staying enrolled, making good progress, and graduating on time. In addition, we want to know if students are transferring to other institutions and to which institutions. We take snapshots of our data to see the percentage of our students who are still enrolled, have transferred, graduated, or dropped out by certain time periods, such as one year, four years, and six years. While tables showing those percentages, such as the one shown in Display 9, are fine for an overall picture, they do not let us dig into the data. However, with the dynamic capabilities of Visual Analytics, you can look at the persistence of any number of groups and see patterns in the data that are hard to analyze looking at each cell in a table, one at a time.
Display 9: Persistence of First-time, Full-time Baccalaureate Students at WKU

Using a report like the one shown in Display 10 allows us to track 4-, 5-, and 6-year graduation rates for various cohorts. We can look at first time, first-year students or transfer students. We can filter by degree level (associate or baccalaureate) and enrollment status (full-time or part-time). We can even track metrics such as underrepresented minority status and low-income status, both of which are metrics in our statewide performance-funding model.

Display 10: Graduation Rates for First-time, Full-time Baccalaureate Underrepresented Minorities

While dynamic snapshots tell one story, a much more interesting story is how students move through the persistence pipeline. Do students drop out and re-enroll often or do they tend to drop out and stay out of school? Do a large proportion of students transfer out and if so, do they go to two-year or four-year institutions? Another great visualization to track the flow of students is a Sankey Diagram, which is currently available using the SAS Visual Analytics Explorer. For a detailed explanation on developing Sankey diagrams to track student persistence, see Helbig (2016).
Display 11: Persistence for All WKU FTFY Students Entering Fall 2009

Display 11 shows the flow of all WKU first time, first-year students who first enrolled at WKU in fall 2009. The first vertical bar shows the students in the cohort. The following six sets of vertical bars represent where the students went after 1, 2, 3, 4, 5 and 6 years. Notice that once students drop out, very few re-enroll at WKU. You will also see that a fair number of students graduate after four years, another large group graduate after five years, but very few graduate between the fifth and sixth year.

Now, look at Display 12, which shows the persistence for students who began their studies in our University College in fall 2009. What do you notice?

Display 12: Persistence for University College FTFY Students Entering Fall 2009

First, many of these students fell into the “Unknown” pipe, which means they dropped out of WKU and did not enroll at another institution. Second, many students transferred out, particularly to two-year institutions. However, the most concerning trend of all was that very few of these students ever graduated.
Most of these students were underprepared—they had low ACT scores and low high school GPAs. When these students enrolled at WKU, our University College was a community college within our University. We had a culture that embraced access, but we just did not have what it took to help these students succeed. Based on persistence models we developed using SAS® Enterprise Miner™, in 2013 we changed our admissions criteria to deny admission to the students who had almost no chance of being successful at WKU. The first year after we changed our admissions criteria, our one-year retention rate of first time, first-year students increased six percentage points.

DEGREE HOURS EARNED

While looking at trends is helpful, the most exciting place to use data visualization to tell a story is when you need to understand an issue and the driving forces behind it. We recently had the opportunity to do that when our Provost asked us to help his Executive Team look into the average number of hours our baccalaureate degree recipients had earned at graduation, a metric in our state’s strategic agenda. The WKU average and the statewide average were both around 139 hours, while the goal was 135. If you knew that the minimum number of hours a student must have to earn a baccalaureate degree in Kentucky is 120, you might logically suspect that most of our students were graduating with an extra 19 hours. However, by visualizing the data, you can see that is not the case.

The first thing we did was to look at a histogram of the data (See Display 13.) The distribution was highly skewed to the right, with no students earning fewer than 120 hours. The median was 135 hours, while the mean was 139, the difference stemming from the handful of students with extremely high (over 180) hours and the fact that no students earned fewer than 120 hours.

Display 13: Total Degree Hours for WKU Baccalaureate Degree Recipients 2012/13 - 2014/15

To investigate the factors contributing to high mean hours earned, we developed tree maps to show the relative number of students and average earned hours for different groups of students. Display 14 shows a tree map we developed using a hierarchy for college, department, and major. As the user clicks on a specific college (HH in the example), the tree map drills down to the departments in that college, then if
the user clicks on a specific department (NURS in the example), the tree map drills down to show the majors in that department.

Just looking at the graphs in Display 14, we noticed an issue with our data. The Nursing (RN to BSN) major is for students who have already earned a degree and are seeking a Bachelor in Nursing degree. By definition, students who had already earned a degree were not supposed to be included in the metric calculation, but they were and were driving our mean degree hours up.

Most of the tree maps we developed showed what we expected to see. For example, transferring a high number of hours, pursuing a major that required more than 120 hours, and pursuing multiple majors or degrees were all associated with a high average number of degree hours at graduation (See Display 15.)

Display 14: Earned Degree Hour Tree Maps with College, Department & Major Hierarchy

Display 15. Tree Maps for Transfer Hours, Degree Requirements & Multiple Majors/Degrees
However, one of the tree maps quickly pointed out a flaw in our logic when we pulled the data for the project. When we looked at the relationship between the number of developmental hours a student took at WKU and the average degree hours, we noticed that students who took no developmental hours had a higher average number of degree hours than did students who had taken 1-6 hours (See Display 16). This did not make sense because developmental hours are courses that do not count toward a student’s degree, but completion is required before the student can enroll in college-level courses and the developmental hours are included in the hours earned at graduation. After some thought, we realized students might take developmental courses elsewhere, then transfer them to WKU, and our current measure of developmental hours did not capture those students. Once we changed our measure to include developmental hours taken anywhere, the tree map made much more sense (See Display 17.) We found it interesting that the number of developmental hours taken did not seem to increase average degree hours until students took seven or more developmental hours.

To allow the Provost’s Executive team to examine a number of scenarios, we developed bar charts that showed the number of students and the average hours above/below the university mean of 139 for the selected students (See Display 18). We provided a number of list controls to let the users see how including/excluding groups of students contributed to the average degree hours. One fascinating scenario, the “best-case scenario” shown in Display 18, selects all of the values that were associated with the lowest average number of degree hours. What was interesting about this scenario was not that it lowered the average hours by a large amount, because it did not, but that so few of our students had no risk factors associated with higher degree hours. Only two percent of our students fell into the “No Risk” category.

One theme that kept coming up as we analyzed the data was the influence the number of transfer hours had on the average number of degree hours at graduation. While this is not a surprising effect because we know that not all hours students transfer to an institution are able to be counted toward degree requirements, we wanted to investigate it further. However, to do this we had to figure out a way to look at the transfer work, without examining over 6,000 individual student transcripts. We decided to use word clouds for the analysis. We extracted all of the transfer course data and the equivalent WKU courses so we could analyze the relationships between the courses transferred to WKU and the equivalent WKU
courses into which they were articulated. We developed two word clouds, one showing the transfer course title and one showing the articulated WKU course. We then created an interaction between the two word clouds where the user could select a WKU course to see what transfer courses were transferring in as that course. (See Display 19.) Although not shown, we also produced a similar set of word clouds, but reversed the interaction so a user could select a particular transfer course to see the associated WKU courses articulated for that course.

Display 18: Best-Case Scenario

Display 19: Transfer Course Word Clouds for Advanced Placement Courses
Display 19 shows the various types of course work included in transfer hours, not all of which are what one would typically consider transfer work. Course work earned through standardized test scores, such as the ACT and SAT, Advanced Placement (AP) courses taken in high school, and CLEP and departmental tests were all included as “transfer hours.” Other types of transfer work included transfer work from the 16 institutions in the Kentucky Community and Technical College System and from all other institutions, both divided into developmental and non-developmental course work. We also included WKU courses awarded for military courses.

Display 19 shows the AP courses taken by high school students and the equivalent WKU course awarded for the AP courses. The first thing we noticed about the word cloud was that the most frequent WKU course for which AP credit was awarded was HIST EL-L, which stands for a lower-level elective in history. In Display 20 we have selected the HIST EL-L course in the visualization to allow us to see the AP courses that transferred in as HIST EL-L, almost all of which were AP American History.

This finding was a little concerning because if you look back at Display 19, you will notice that the most frequently articulated AP course was American History, which meant that a large number of students were taking AP American History courses and getting only elective credit for them. While elective credit is better than no credit, it is not as good as credit for a course that meets degree requirements. One of the individuals with whom we shared the word clouds asked why this was the case since we advertised that we awarded college credit for HIST 240 (US History through 1865) or HIST 241 (US History since 1865) for AP History, which could clearly be seen on our website (https://www.wku.edu/registrar/documents/form_creditbyexam.pdf). We researched this further and discovered that WKU started awarding credit for HIST 240/241 in summer 2013. Prior to that, WKU awarded only elective credit. Since the analysis used data for students who graduated from WKU...
between 2012 and 2015, very few of the students would have had high school AP coursework accepted in 2013.

Another discovery we found using the word clouds dealt with developmental coursework received from the Kentucky Community and Technical College System (KCTCS). As shown in Display 21, the WKU course most frequently awarded for developmental work from KCTCS was DEVL EL-L, which is a lower level developmental elective. In essence, these students got credit for nothing, because developmental electives do not count toward anything—they do not meet degree requirements or even developmental requirements, if the student needed to take developmental courses prior to taking college-level courses. Display 21 also shows the KCTCS course most frequently transferred in as a developmental elective was Basic Algebra with Measurement. A check of our statewide website that shows what each course at each Kentucky institution transfers to at other Kentucky institutions (http://knowhow2transfer.org) shows very clearly that we award elective developmental credit for Basic Algebra with Measurement. The question remains whether or not students truly understand when they take the course that it will not count toward anything at WKU.

Display 21: Developmental Coursework Transferred from KCTCS to WKU

One last finding the word clouds revealed involved credit awarded for military courses. Once again, the most frequently awarded credit was for a lower-level elective, TRAN EL-L (See Display 22). The military course most frequently awarded elective credit at WKU was Basic Combat Training. This course, like many other military courses, does not have an equivalent WKU course, so the only credit we can award is elective credit. While visualizing the data in multiple ways, we noticed that although there were few students who transferred in military credit, when they did transfer it in, they transferred in a large number of hours, many of which we can see came in as elective credit. This is indicative of the reason why high
transfer hours is associated with high degree hours—many courses students transfer into WKU do not have a WKU equivalent, and as such, transfer in as electives, usually at the lower level. Most degree programs count very few elective hours, and most of the required elective hours are at the upper level. Therefore, students transferring in a high number of hours will likely not get to count many of those hours. Students transferring in military hours tend to transfer in a large number of hours, most of which do not count toward degree or general education coursework, so they graduate with a large number of hours.

CONCLUSION

From Sankey diagrams to tree maps and word clouds, SAS Visual Analytics provides you with the tools necessary to find and tell the stories in your data, discoverable only through the pictures provided by data visualization. By allowing a diverse group of users access to the data in visual formats tailored to meet their needs, each user can discover stories they may or may not have expected to read. As data scientists and developers, data visualizations also take your understanding of the data and its nuances to another level. Only the size of your imagination can limit the stories you can uncover through data visualization. All you have to do is get started with your mouse.

REFERENCES


RECOMMENDED READING

- *SAS® Visual Analytics 7.3 User’s Guide*


CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at

Tuesdi Helbig, Ph.D.
Director of Institutional Research
Western Kentucky University
270-745-3250
Tuesdi.helbig@wku.edu
https://www.wku.edu/instres