CROSSTOWN TRAFFIC AND BEYOND

Jutta Hoegmander

Master of Arts in Specialized Journalism

University of Southern California

August 2017

CONTENTS

1 ABSTRACT

Crosstown Traffic started as an ambitious project, the mission of which was to combine the skillsets of both journalism and computer science students to produce data-driven journalism that would deepen, broaden and demystify the public's understanding of traffic in Los Angeles. It was the first real collaboration between Annenberg School of Journalism and Viterbi School of Engineering at the University of Southern California.

In my thesis, I explore how this ambitious project could continue so that its impact extends beyond the calendar of the semester. This includes trying to reconcile the differences between the cultures and methods of working of journalists and computer scientists. In addition, I will try to describe future projects and how they could be incorporated into the work flow of USC Annenberg. I use my own experiences about taking part in the collaboration and the discussions I have had with different stakeholders as the basis for my suggestions.

Though I am focusing on the Crosstown Traffic example, the implications this thesis attempts to raise are much broader. Data journalism is an increasingly powerful and widespread method of communication. Yet it requires resources and skills that are often in short supply in today's newsrooms. Examining how a single project can extend its lifespan and usefulness could provide lessons for how to get more returns on investments in data journalism. Studying interactions between data experts and journalists can also reveal ways to make these types of collaborations more common and useful.

2 INTRODUCTION

This is the story of Crosstown Traffic, the first joint venture of the Annenberg School of Journalism and the Integrated Media Systems Center of the Viterbi School of Engineering at the University of Southern California. It brought together students and faculty from two different fields: journalism and computer science. The aim was to combine the strengths of both institutions to produce original journalistic content based on rigorous data analysis; in other words, data journalism. This would form the basis for new, more informed conversation around public policy and transportation.

There is hardly any topic people in Los Angeles care more about than transportation. It is part of the everyday fabric of the city itself. The issue is now more pressing than ever because 71 percent of the voters in Los Angeles County decided to vote in favor of Measure M in November 2016. It meant that the vast majority of voters were prepared to raise their sales tax indefinitely to support the building of new transit projects by an estimated sum of $120 billion over four decades (LA TIMES). That makes producing comprehensive information on transportation issues extremely important.

There has been an information void about the real state of traffic in LA. Crosstown Traffic has the potential to provide Angelenos with the means of keeping themselves informed about the past and current conditions of the infrastructure in addition to being able to see possible trends and changes in its performance over time. Thus, it facilitates informed discussion and engagement between the voters, transportation officials and policymakers.

All of those mentioned want to solve the traffic problems in Los Angeles; they just have not had the means to communicate with each other on the same level. There are commercial,

personal and political interests involved that can distort the communication. A university such as USC has the ability to act as an unbiased mediator between these different interests and give them tools that help them see the problems and the potential solutions in a new way.

There is great potential in data journalism at all levels of reporting but it could be especially valuable for local and regional coverage (Ali 2014). For USC Annenberg, this is just the first stab at trying to employ data-driven journalism in a way that encourages civic engagement and more informed discussion and decision making. If databases and dashboards are maintained and updated regularly, they can be a valuable resource for all stakeholders. These new methods of providing information in an interesting way are crucial for the media industry currently struggling to legitimize its place in the everyday lives of people.

I became involved in Crosstown Traffic as a part of Prof. Gabriel Kahn's directed research class during fall semester 2016. As a professional journalist with over ten years of experience in news reporting, I recognize news when I see it. I became interested in the project immediately when I heard that the Integrated Media Systems Center (IMSC) of Viterbi School of Engineering had access to the Los Angeles County Metropolitan Transportation Authority's (Metro) databases on traffic accidents, the speed and volume of cars and the location data of the LA Metro buses and that we would be writing data journalistic stories based on that data.

I decided to craft my thesis around the project. The thesis consists of four parts. The first part are the two data-driven stories I wrote during the fall semester in collaboration with the Viterbi engineering students. In the beginning of the spring semester 2017, I began the second part of my thesis and prepared all three dashboards for publication. The work consisted of

copyediting and trying to make the user interfaces more user-friendly. This was done in collaboration with the students from Viterbi, Professor Kahn and other Annenberg students.

The stories and dashboards were published as a part of the Crosstown Traffic project by Annenberg Media on March 21, 2017. The historical project gathered the attention of several news outlets at the time of the launch. It was covered by The New York Times' California Today newsletter, local TV stations ABC, NBC and CBS, Business Insider and KCRW among others.

At the same time, I began to look ahead and devise how the data could be transformed into a package of easily accessible and significant data that could be published monthly. The third part of my thesis is a plan for that future dashboard.

The collaboration between journalists and computer scientists can produce results none of the parties involved can yet fathom. In addition to producing high-quality and high-impact data journalism, forces could, for example, be united to solve problems journalists face in newsrooms all around or more specifically, at the Annenberg Media Center. By giving both fields the chance to use their own strengths in sync with the other, USC is leveraging both. For me personally, Crosstown Traffic has been an exhilarating demonstration of the power of collaboration and utilization of the different skill sets of those involved. My thesis is one step in trying to make sure that this project will continue to live on as its own living and breathing ecosystem. As the fourth part of my thesis, I will concentrate on the lessons that can be learnt from Crosstown Traffic in order to make future cooperation between USC Annenberg and USC Viterbi even more effective.

3 THE BIG PICTURE: AN INDUSTRY IN CRISIS

The digitization of the media industry has brought data into newsrooms in a way never seen before. Data pervades nearly every step of the production and consumption of journalism. Media outlets use data and analytics tools to follow the performance of individual stories to track what their audience is interested in and adjust the production of content accordingly. At the same time, modern tools make monitoring the behavior of their audience simple. Audience engagement professionals try to make sure that the stories perform as well as possible. (Cherubini 2015)

All this ties into a broader picture of the changes taking place in the media industry. News consumption keeps shifting to online and mobile platforms (Matsa and Lu 2016). Competition for eyeballs in the digital realm is fierce; there is an abundance of media content available to choose from, and big players Facebook and Google dominate the digital advertising space. This leaves traditional publishers and cable companies with declining print subscription and advertising revenues and a strong desire to go digital. The focus turns to the audience and their needs. (Thompson 2015) Traditional legacy media companies, such as The New York Times, are now trying to target and engage with their audience and find captivating, digital ways of presenting stories (Leonhardt et al. 2017).

I, myself have witnessed a complete transformation in the working culture of the newsrooms from the year 2000 and my first summer internship to today. Nowadays, even individual journalists see their stories as products they have to “sell” to their audience by presenting them in the most attractive possible means. This is a big shift from the past, when journalists could see themselves as the only ones setting the agenda and deciding what the audience should be interested in with little regards to how well the audience received the final product. As I have

learned during my studies on emerging and mobile platforms at USC Annenberg, user experience becomes even more of a priority when journalism is made mobile in mind. Mobile phones are a very personal means of communication and invading that space with, for example, annoying push notifications can make people turn those notifications off.

The new digital media space needs new forms of storytelling. For long, publishers just transferred their print content online without using the properties inherent to modern day Internet: interactivity and the ability to combine text, photos, video and audio in compelling ways. What works in print, doesn't work online or on mobile. New platforms, such as news apps and social media, use their own conventions of telling stories.

Some have turned to data journalism as a new, powerful way of communicating stories. Data journalism uses the vast amount of digital information available online and tries to turn it into newsworthy stories. Data visualizations bring the numbers to life and show relations between things more effectively than a mere description in a text piece ever could. Interactive data visualizations can give users the ability to play with the data and focus on those details they are most interested in. (Bradshaw 2012) This is an example of mass customization, the idea that a product can be customizable and have different meanings and unique value for different individuals even though it is produced efficiently for masses (Gilmore and Pine 2016). This was also a goal for the dashboards of Crosstown Traffic. Traffic as a subject lends itself extremely well to mass customization because it happens to involve everyone.

All the mentioned revolutions make collaboration between journalists and computer scientists more and more frequent and necessary in today's newsrooms. Simultaneously, the ideal skillset of a journalist is evolving. For example, USC Annenberg aims to educate tech- and

data-savvy all-round journalists ready for the challenges of today's journalism. All journalism students earning a Master of Science degree are required to take classes in data journalism and coding.

Journalists still desire to tell good, impactful stories even though the means of doing that are changing.

4 CROSSTOWN TRAFFIC: THE STORIES

From the very beginning of Crosstown Traffic, we knew that the datasets we had at hand were packed with stories that just waited to be told. These datasets contained unique information no journalist had had the chance to report on before; in other words: news. The challenge turned out to be choosing which stories to work on. The decisions were made based on the perceived importance and impact of the stories in addition to an assessment of whether tackling the stories was possible from the data analysis perspective in the amount of time we had. To sum up, we were thinking about the data in terms of turning it into newsy headlines. And we got our headlines, including the most accident-prone roadways, the most hazardous travel hours, the worst interchanges and the most delayed bus route.

The goal was also to transform the data into stories that would be real and relatable to people living in Los Angeles. As in all data journalism, we wanted to see behind the numbers, see trends and changes and how they affected the lives of people. Using a person and their experiences to illustrate the numbers is a common and powerful way to tell data-driven stories. (Mitchell 2015)

At the same time, the Viterbi students and faculty were engaged in a complex data science process. During the data extraction stage, they tried to determine which data intervals were the most useful for the project and spent weeks trying to define which months and years contained the best data. After that, the data was cleaned with statistical measures so that, for example, outliers did not affect the results. Data was normalized to remove redundant data. Then the data was analyzed according to the definitions we agreed on. The final stage was visualizing the data in the dashboards.

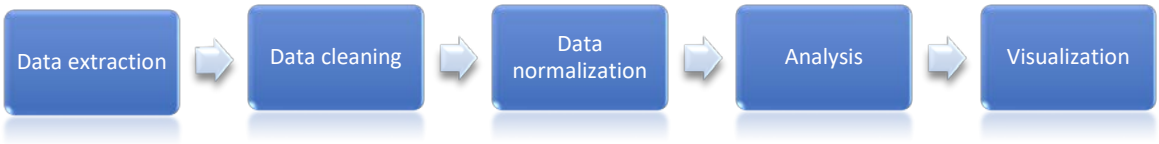

Fig. 1. The data science process

The Annenberg students taking part in professor Kahn's directed research class were assigned to three categories of data: the accident group, the congestion group and the public transportation group. Each group had a counterpart from Viterbi who was responsible for the data cleaning and analysis. Each Annenberg student was expected to craft one story based on the findings from the data and take part in designing a dashboard for the group at hand. Though data-driven, the stories turned out to be traditional text pieces with the exception of one video designed for social media.

I was part of the group that focused on public transportation in Los Angeles, specifically the LA Metro buses. The dataset was complex and needed a lot of cleaning and verifying work. It consisted of location data provided by GPS transmitters attached to LA Metro buses and had 16 million data points. The IMSC wanted to be sure that the data we based our stories on would be valid. At some point, it was uncertain whether the data analysis would be finished early enough for us to have enough time for reporting and writing the stories. There were some story ideas we were interested in that could not be executed because of that. The analysis was based on data gathered from October 1, 2015 to September 30, 2016 because the dataset from that period was complete and reflected the possible seasonal variations in the transportation system. We decided to concentrate on determining how on time the buses were. The computer science students were

able to analyze that by comparing the location and time data of the GPS transmitters to those of the official schedule.

Based on the analysis of the data, I wrote two text pieces that comprise the first part of my thesis. One was a story of one of the most delayed bus lines, the 720. The second story was an overview of the most delayed LA Metro bus lines and the reasons behind the poor on-time performance of those buses.

4.1 THE TOP FIVE MOST DELAYED BUS ROUTES

When trying to figure out which stories to report on, one aim was to grab the low-hanging fruit from the data: those stories that would be newsworthy and interesting no matter what the results of the analysis were. Listing the most delayed bus routes was one of those stories.

Initially, I had interviewed Bob Holland, the Sr. Executive Officer of Bus Operations at Metro Los Angeles, so that I could incorporate his comments to my main story about the 720 bus route. During the reporting process, I decided to divide the stories so that the main story would concentrate on that specific route and the experiences of the people riding the buses. LA Metro's explanations and the top five most delayed bus routes comprise the other, shorter story. There was one, hardly surprising common denominator to the poor on-time performance of the routes: traffic.

You can find the transcript of the story as an appendix of this thesis or use the following link: http://crosstowntraffic.uscannenbergmedia.com/genericjutta2.html

4.2 Riding One of the Most Delayed Bus Routes

Finding the most interesting story to write from the public transportation data was a difficult task for me. I spent hours combing the data and trying to find meaningful connections between different stats. Finally, I noticed that one of the most delayed bus routes travels between two of the most delayed areas: East Los Angeles and Santa Monica. My working hypothesis was that nannies and other lower-income workers would use the bus to travel from East LA to Santa Monica and back. I didn't manage to find my nanny. While riding the 20-mile-long route twice back and forth, I also realized that most people don't take the whole route but jump off somewhere in between. I decided to focus on how the lateness of buses affects the residents of Santa Monica and East LA differently because of the demographical differences of the areas. People of lower income level rely more on public transportation than those who are financially better off (Taylor 2016).

You can find the transcript of the story as an appendix of this thesis or use the following link: http://crosstowntraffic.uscannenbergmedia.com/genericjutta.html

5 CROSSTOWN TRAFFIC: THE DASHBOARDS

Digital storytelling gives journalism the possibility to think outside the standard article "box" and experiment with more interactive and visual means of communicating with the audience (Bartlett 2013). That encompasses data visualizations as a powerful way to illustrate the numbers produced by data-driven journalism. The data visualization can be a complete story in itself in a way that doesn't require a traditional text piece attached to it. As expressed by Aron Pilhofer, the former Interactive News Editor for the New York Times: "There are times when data can tell a story better than words or photos." (Pilhofer 2012)

There are great examples of visualizing databases in a way that relates to earlier discussion about mass customization. The Texas Tribune built a government salary explorer that currently holds the salaries of 500,000 state and municipal employees in Texas. It gives users the opportunity to play with the data. You can customize the search by typing in the last name or title of the employee or by choosing to search by government entity. The Los Angeles Times illustrated the differences between street quality grades in Los Angeles by making an interactive map with different color assigned for each grade. By typing in an address, you would get a close-up of the map that would show the grade of that street and its neighboring streets. (Schleuss and Poston 2016)

We wanted to follow this path and experiment with building traffic dashboards that would allow the users to interact with the data and drill down to the specific details that interested them and affected their lives. That way we would be visualizing the data in a way that would be relevant for people. If a person uses the 405 every day to drive to work, they will probably be most interested in learning about the congestion on that particular stretch of freeway. The

dashboard should still give the user the possibility to check the situation on other freeways, too. This customization might help to fill the information void that blocked people from being knowledgeable enough to discuss the state of traffic in Los Angeles in an informed way.

To give all the stakeholders, including not just the public but also policymakers, politicians, researchers and journalists, the possibility to find answers to their questions, three different dashboards were included in Crosstown Traffic: one for each group.

The main problem with the dashboards was the fact that they had initially been designed by the Viterbi students to be mock-up versions of the dashboards. With the help of these mock-ups, we, the journalism students could see what the dashboards were like and what kind of data they could include. The dashboards also acted as user-interfaces to the data for us journalism students so that we could find data relevant for our stories there. Thus, the dashboards needed a fair amount of polishing before they could be published to a wider audience. The dashboards could have turned out even better if a UI and data visualization specialist would have been there from the very beginning to guide the designing of the dashboards.

I copyedited all the texts of the dashboards with the help of professor Kahn, who had the final say in all editing decisions. The texts were formulated so that they would serve both the average reader with no knowledge of data analysis and a researcher who was interested in how IMSC had executed the data analysis. I tried not to oversimplify the explanations so that the essence of the work would not disappear.

A big part of the process was making the dashboards more user-friendly. That was done by inserting pop-up boxes that told the users what the dashboard was all about and how they could

interact with it. In addition, emphasis was put upon crafting informative headings and labeling things so that the dashboards would be as easy and intuitive to use as possible.

Design issues, like font and color scheme, were also addressed and changed so that all the dashboards had a uniform look.

5.1 How On Time Are LA Buses?

The public transit dashboard was designed to illustrate how on time or delayed the LA Metro buses were in relation to their schedule. The aim was to both portray the overall situation and to let people play with the data. To give people the possibility to see the on-time performance of the buses they used or were otherwise interested in, we created several interactive visualizations that probed the issue from different viewpoints.

The first was a map that included the list of all the bus routes in the county sorted from the worst performers to the best or the other way round. By clicking on a route number, the user would be able to see the route on the map and more details about its performance. Thus, users could try to search for common patterns in the best and worst routes or focus on the routes that were meaningful to them.

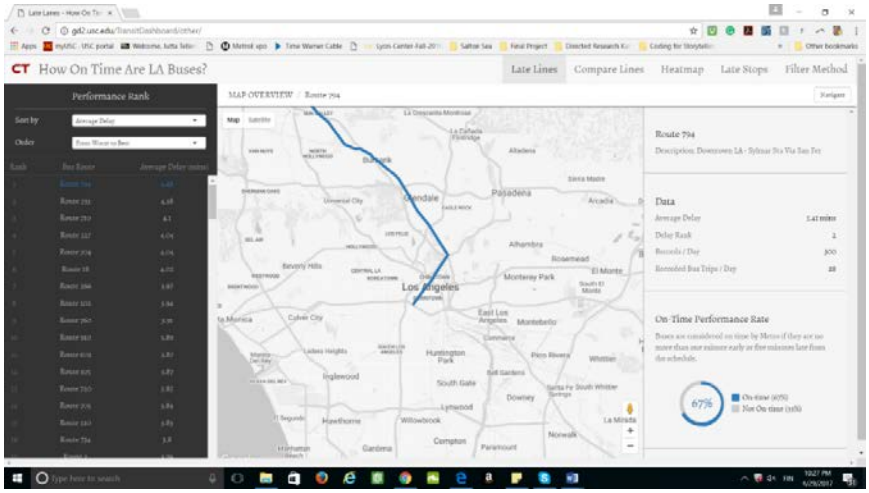

Fig. 2. Late Lines gave users the chance to see details of the most and least delayed bus lines and their route on a map.

The second interactive visualization was designed to let people compare the performance of two bus routes with eachother. That way they could, for example, see if the Metro Rapid lines were actually faster than the local lines as they were supposed to be. The third interactive was a map of the most delayed bus stops in the county. To give an overview of how delayed the buses were in Los Angeles during different hours of day, we also included a delay heatmap that highlighted the on-time performance with different colors.

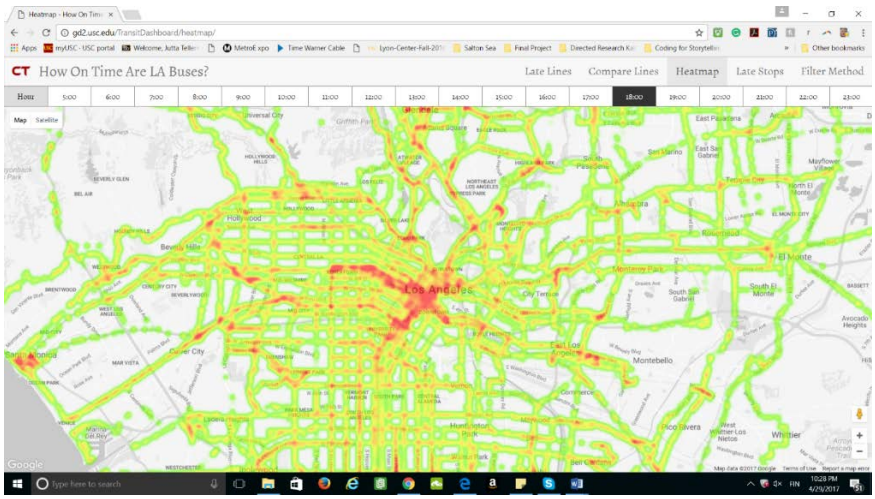

Fig. 3. The Heatmap showed how the on-time performance changes during the day.

Link to the public transportation dashboard: http://gd2.usc.edu/TransitDashboard/

5.2 Accidents Across Los Angeles

The accident dataset consisted of accident reports from a period of four years so it was possible to try to discover overall changes and patterns, such as had the number of accidents gone up or down and what was the most dangerous day and hour of the day to drive a car in Los Angeles.

The data visualizations of the accident dashboard allowed the users to reflect their own experiences on what the results of the data analysis revealed. Thus, the data was made more relevant for them and their lives. An interactive visualization provided the users of the dashboard with a chance to play with the data to see how the number of accidents changed in relation to different time parameters such as month and day of the week. Another let you compare years to each other.

Fig. 4. The Accident Timeline portrayed the number of accidents by day of the week and hour of the day plus included a heatmap of when accidents had happened.

To give a better understanding of where accidents were happening, the accidents were portrayed on a map. In addition, the dashboard included static images of the top 10 most collision-prone places on freeways and surface streets by year.

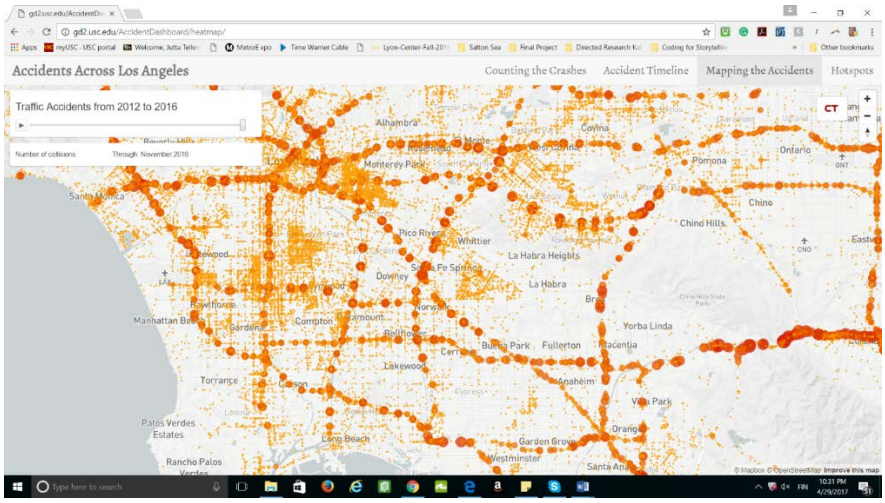

Fig. 5. Mapping the Accidents demonstrated where the accidents had happened.

Link to the accident dashboard: http://gd2.usc.edu/AccidentDashboard/

5.3 How Fast Is That Freeway?

The congestion group had at its hands one of the most burning questions concerning traffic in Los Angeles: has the traffic gotten worse over time? Choosing the right variables from the available data to illustrate that question took its time.

A measure tangible enough for everyone who had tried to get from place A to place B with a car in LA was found in average speed on different highways during rush hour. First, the congestion group had to determine what was considered as rush hour. The interactive visualization let users choose a freeway and then compare speeds during morning and evening

rush hours to those of previous years. That way they could also see which places were the most congested along that highway.

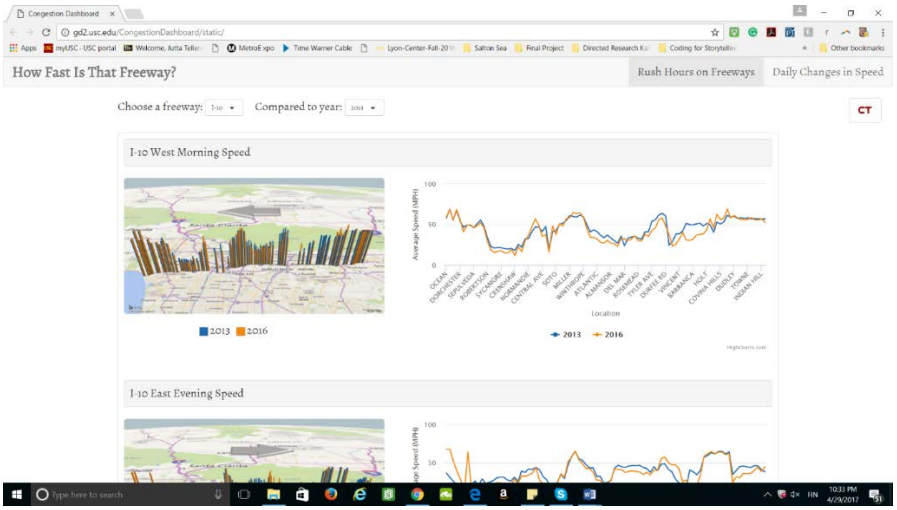

Fig. 6. Rush Hours on Freeways prompted users to pick a freeway and compare its rush hour speeds in 2016 to those of previous years.

Another visualization focused on daily changes in speeds along different highways.

Link to the congestion dashboard: http://gd2.usc.edu/CongestionDashboard/static/

6 Analysis of the Collaboration: Fixing the Potholes

The need for journalists and computer scientists to collaborate will continue to grow as the digitization of media deepens. Most journalists are not equipped with skills that allow them to perform complicated data analysis or build mobile apps and sophisticated multimedia presentations from the scratch. They need help. To make this kind of cross-cultural team work easier in the future, it is crucial to take the time to analyze what kind of problems we faced during Crosstown Traffic, what could have been done differently and why. By identifying the potholes on the road, we can try to fix them in order to have a smoother ride to the destination next time. This might have larger implications for future students taking part in collaborative projects. Learning how to "talk engineer" is an important skill in the newsrooms of the future. It is just as important for the computer science students to understand where journalists are coming from. The media industry will be one possible employment option for Viterbi students taking part in collaborative projects with Annenberg.

The basis for fruitful collaboration between Viterbi and Annenberg is solid. Viterbi students have the technical skills and the Annenberg students the journalistic craftsmanship to be able to produce quality data-driven journalism together. The different skillsets and mindsets are an asset if they are utilized in the best possible way. However, the differences might also lead to confusion and inefficiency if not addressed properly.

According to Emily Bell, the Director of the Tow Center for Digital Journalism, "there is always a right answer in engineering, whereas in journalism, there are only more questions." Bell portrays the cultures of tech and engineering as fundamentally different from each other. She goes on to describe how the tech people at the Guardian longed for precision and certainty while

journalists seemed to enjoy ambiguity and uncertainty. (Bell 2016) However, there are tech-minded journalists around as well as engineers who have a passion for journalism. The perfect journalist of the future might be a “Digital Media Data Guru”, who would embody the best of both worlds (Schmitz Weiss and Royal 2013).

The first step in bridging the culture gap is identifying the possible problems and acknowledging the differences in cultures without exaggerating them.

The main problems during Crosstown Traffic were related to project management and the interaction between the engineering students and journalism students.

In the future, professors leading the collaboration could act as mediators between the two worlds. That includes describing the expectations of both parties and their different methods of working. Both journalism and computer science share an interest in information but computer scientists are used to well-defined projects and systematic thinking, whereas journalists are more prone to just go with the flow (Codrea-Rado 2012). In data-driven projects, journalists are searching for the story, news, a change or a trend without necessarily knowing in advance what it is and where they will find it. That might be a problem for the computer scientist, who wants to have a clear-cut problem he can focus on solving.

A more rigorous approach to project management could make collaboration easier. That would include a task list that would incorporate the expectations of both parties and clear deadlines for completing the tasks and the whole project. If surprises or delays would occur, the list would be revised accordingly and if time would run out, tasks would be prioritized. For example, the computer science students were waiting for requirements for the dashboards but we never delivered them because we didn’t know they were needed and actually didn’t know what

we wanted from the dashboards at that stage. Our counterparts from Viterbi solved the problem creatively by building mock-up versions of the dashboards so that we could see what they could be like.

In addition, the launch of the project was delayed multiple times during the spring semester. It might have been wise to start asking questions from the datasets earlier on. That way the validation and verification of the data could have been finished earlier. Now the last checks to the data were made during the spring semester when most of the stories were already finished. If there would have been major changes to the data the stories were based on, it could have meant complete restructuring and rewriting of the stories. Having a clear schedule and deadline would have made the process more effective. That would have also helped students to prioritize and allocate their time better.

The whole process would have been more efficient if the two parties would have been more closely in contact. That would have made real team work possible. Now both parties concentrated on their own tasks separately and precious time was wasted on trying to figure out things that could have been dealt with easily together. It would have made a difference if the Viterbi students would have been officially part of the directed research class or at least required to be present in class regularly. Now we saw each other rarely. Face-to-face interactions were fruitful because everyone was able to ask follow-up questions right away if they did not understand what the other was trying to convey.

In addition, there was a knowledge gap between the students that hindered the interaction. The computer science students were executing a difficult data science process the journalism students did not fully understand. We as journalism students were seeking to write good stories

and the computer science students tried to provide us with the required data. We formulated questions in layman's terms; they had to translate them to variables that could be measured and actually found in the datasets. At times, it felt like we were speaking a different language.

The lack of team work skewed one of the most important stages of the whole project: trying to figure out which questions to ask from the data. First, we, the journalism students came up with questions we were interested in. Those questions were compiled into a list. Then the list was presented to the engineering students who told which questions were feasible and which not. A quicker and more efficient way would have been to come up with the questions in collaboration with the engineering students who knew the limitations and possibilities of the datasets. During later discussions, they came up with good questions that would have been great stories but it was already too late to start working on them as a part of the class.

In general, a more systematic approach in trying to figure out which questions to ask from the data might have been helpful in finding the best possible story ideas. The end-result was good but the process was random. If we would have employed system approach, we would have first analyzed what data there is, then concentrated on figuring out systematically what objects or concepts could be derived from the data and then formulated the questions based on the objects found interesting from a journalistic perspective. Thus, we could have been more certain that we had come up with the most important questions and stories.

To sum up, there is a clear cultural and methodological gap between journalists and computer scientists but it can be bridged. USC Annenberg is the perfect place for this kind of experimentation because the school is already trying to provide its students with basic skills in coding and data. It also offers classes that focus on, for example, building apps and products. In

collaborative projects, the students will have an amazing opportunity to do ground-breaking work and learn best practices from tech side. Experience in working with computer scientists will be an asset for them once they head into the real world and start working with them.

7 PLAN FOR THE FUTURE: SNAPSHOT OF LA TRAFFIC

Data-driven journalism is not an easy, quick or cheap way of telling journalistic stories. In fact, it is the antidote to fast-paced and fragmentary online news production so pervasive these days. High-quality data journalism is an investment of time and resources. In return, it should have a longer life than your standard daily set of news. If executed wisely, data journalism can be an investment in the future.

A good example of a long-lasting data journalism product can be found in the Los Angeles Times's Mapping L.A. project. It began as an attempt to draw the lines of different neighborhoods of the city of Los Angeles in 2009. It sparked lively conversation and the lines were redrawn with the help of readers. Now the interactive map covers the whole Los Angeles County and contains statistics about the demography, crime, and schools of different cities and neighborhoods. It is still updated regularly and has become an important source of information for many stakeholders. Not all data-driven projects turn out as successful as Mapping L.A. Nevertheless, data visualizations and dashboards should be updated from time to time so that their information value stays high.

In accordance, Crosstown Traffic should not be a one-off occasion but a living, breathing entity that will continue its existence long after the first publication. One possible way to accomplish this is to design a dashboard that would be updated monthly. It would be a monthly snapshot of traffic in Los Angeles that would accumulate into an overview of the status of traffic.

To try to make the continuation of the project as certain as possible, I have used these three premises as my guideline when designing the dashboard:

1) The monthly cleaning of the data should be made as automatic as possible. Thus, the continuation of the dashboard will not depend on people manually doing the work. In addition, the dashboard should not be dependent on the input of any single person and should live on after the graduation of the students involved. It could be executed by the Viterbi students involved in Crosstown Traffic because they know the data and its cleaning process better than anyone. By writing a script that will make the process of gathering and cleaning the data automatic, they will minimize the need for manual labor in updating the dashboard monthly.

2) The dashboard should provide valuable information and be useful to different interest groups, such as the public, the media, LA Metro, and researchers.

3) The monthly updates should be integrated into the workflow of the Annenberg Media Center so that one news editor would be responsible for checking the updated dashboard and seeing if there are newsworthy changes in the data. By assigning at least partial responsibility for the dashboard to someone at the Media Center, we will try to make sure the Media Center will be able to take advantage of the asset they have at their disposal.

The first premise obviously limits the complexity of the dashboard. The transportation datasets used in Crosstown Traffic had to go through multiple rounds of data cleaning before they were ready to be used in the dashboards. The same would not be possible on a monthly basis for an indefinite period of time.

To give people a clear overview of the traffic situation at that point of time, my proposal is a dashboard that would contain the three most vital stats for traffic in Los Angeles: the number of accidents, the on-time performance of the LA Metro buses and the volume of cars. These three stats are currently not readily available online – especially on a monthly basis. This snapshot of

traffic in Los Angeles would also offer users the possibility to track change by comparing the new monthly stat to the previous ones.

Based on discussions with transportation reporter Laura Nelson from The Los Angeles Times, especially the number of accidents and the on-time performance of buses would be of interest to reporters concentrating on transportation issues. The volume of cars could be used as an indication of congestion but also of just changes in the number of cars on the road. These stats could form together the basis for continuing discussion between the policy makers and citizens about the current state of traffic. In a discussion about the future of the project on March 9, 2017, Program Manager Kali Fogel from the Los Angeles County Metropolitan Transportation Authority stated that people want reliability and alternatives when it comes to traffic. This monthly dashboard would give Angelenos the possibility to determine whether hopping on a bus could be an attractive option to driving a car.

To give some idea about the appearance of the dashboard, here are some examples of possible ways of illustrating the data. Below is a visual presentation of the social media mentions of the Financial Times. The monthly number of accidents could be presented with the barometer graph on the left so that the average number of accidents in that particular month would be in the middle of the half circle. If the number of accidents during the month at hand would have been greater than the average, the pointer would be on the right side of the half circle that would be red. The left side of the barometer would be green because the number of accidents would be below the average. The graph on the right would be a good way to show how the number of accidents has changed over time. In order to make the data relevant to the lives of people, it would be good to have a map of the locations of the accidents.

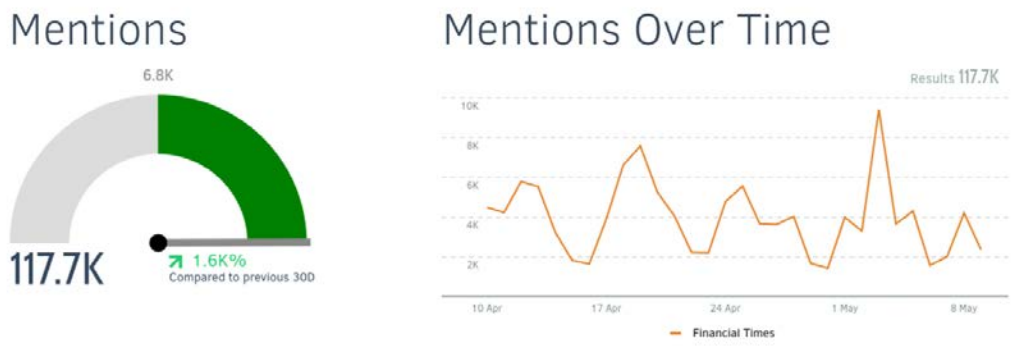

Fig. 7. Social data company Talkwalker's illustration of Financial Times's social media mentions could be transformed into a visualization about the monthly number of accidents.

The same barometer could be used to show whether that month's on-time performance of LA Metro buses would be above or below the average for that month. The dashboard should also include the list of the most delayed bus routes during that month with reference to their previous ranking. Seeing them on a map would give users needed context as to where the delays are happening. In the example below, routes and delayed stops in San Francisco are included in the same interactive visualization.

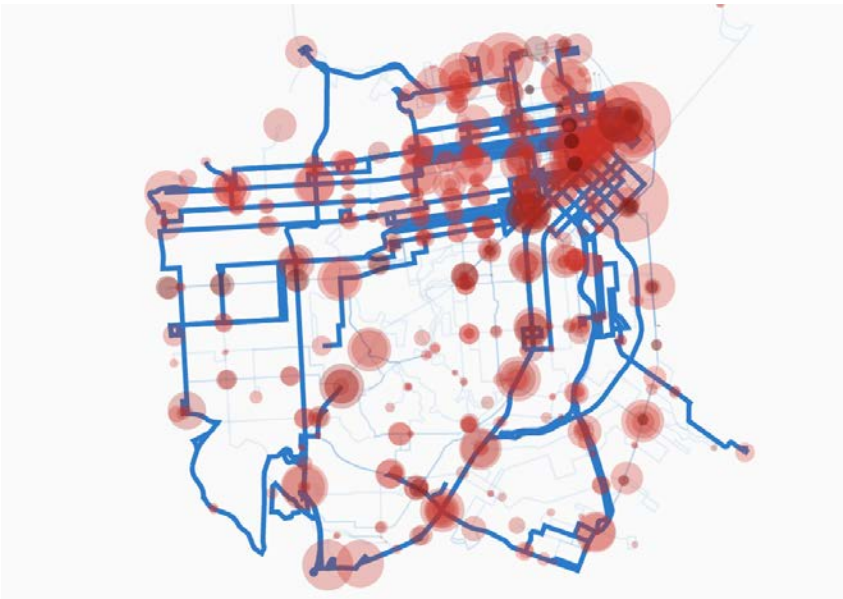

Fig. 8. This award winning illustration about public transit in San Francisco combines the routes and delayed stops in one interactive presentation (Sutedjo-The and Lee 2013).

The barometer could again be used in describing the monthly volume of cars on major freeways and its change in relation to the average volume. This stat could be brought to a more granular level by using a map. The following map of the city of Pilsen in the Czech Republic shows how the volume of cars during a time period of an hour can be illustrated on a map.

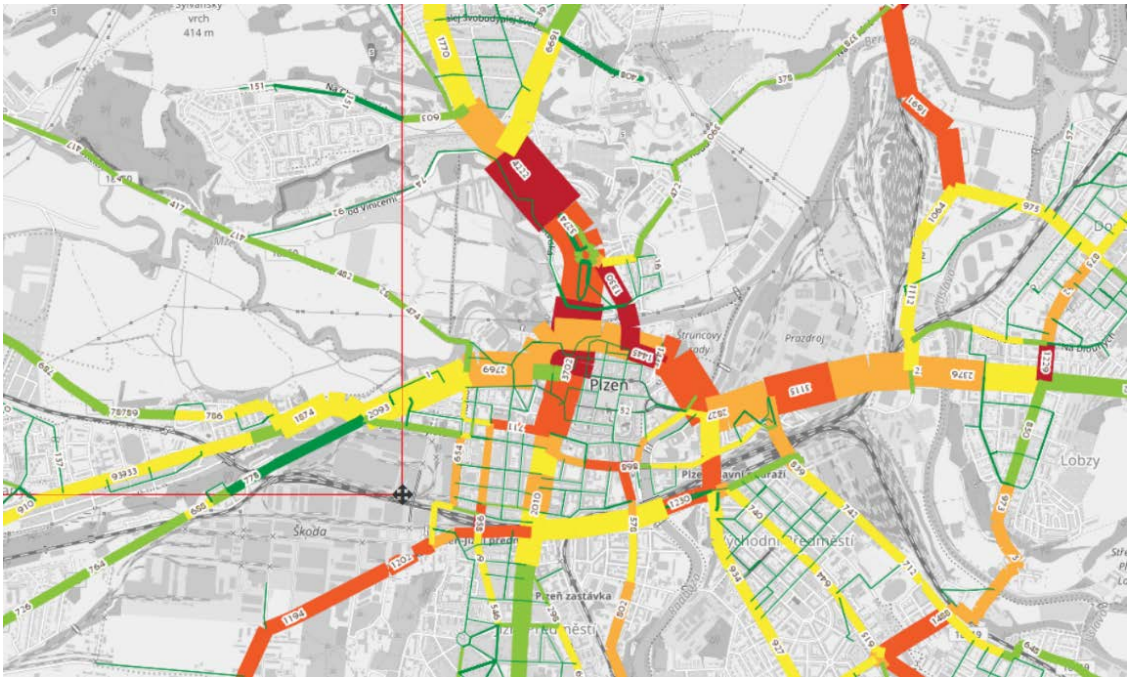

Fig. 9. This map shows how the colors and breadth of lines can illustrate differences in the volume of cars

8 CONCLUSIONS

There is great potential in the collaboration of USC Annenberg School of Journalism and USC Viterbi School of Engineering. Crosstown Traffic can be seen as a model of the way Annenberg can work in cooperation with Viterbi and produce something that is unique and valuable for both parties. This parallels the broader trend of collaboration between tech and journalism taking place in the media industry.

The project will hopefully live on in the future. The monthly updating dashboard will provide different stakeholders with means of communicating about the traffic in Los Angeles. The three broader dashboards can be updated in regular intervals, too. That way Crosstown Traffic will become a continuous resource of information and discussion.

In addition, there were stories that could not be written during the fall 2016 directed research class because the data had to be verified. Those ideas are up for grabs now besides new ones that might arise from the updates to the dashboards. Different datasets can be combined with those of Crosstown Traffic. For example, the effects of weather conditions, such as rain, on LA traffic would be an interesting topic to dig into. Housing prices, air pollution and the state of the economy are just few of the other possibilities.

In Crosstown Traffic, the stories we produced were text-based with the exception of one social video. Thus, there is still plenty of room for exploring new ways of digital storytelling besides the dashboards. In a brainstorming session about the future of Crosstown Traffic at the Annenberg Innovation Lab on April 26, 2017, the idea of incorporating art and possibly the Roski School of Art and Design was introduced. New technologies, such as AR and VR, could also provide new insights into the data. One opportunity would be to project data in the urban

landscape with the help of AR. It could be possible to, for example, project the heatmap of the lateness of buses on the actual streets so that you could judge the lateness on a certain street by the color of the AR projection. A timelapse video about a day of traffic in Los Angeles could be worth exploring. Stylized heatmaps of accidents, on-time performance of buses and the volume of cars could be converted into beautiful animations.

Fig. 10. Uber decided to tackle data visualizations by mapping the major routes their drivers used in different cities. An animated timelapse video of a day of traffic in Los Angeles could borrow elements from this stunning presentation.

There are other possible ways to make use of the newly-formed close acquaintance between Annenberg and Viterbi. For one, the Annenberg Media Center could have a proper data desk that would consist of both computer science students interested in journalism and journalism students interested in data journalism. The use of data in newsrooms will continue to increase so the

students should have the chance to dive into data at the Media Center, too. As an educational media organization, the Annenberg Media Center is free from the economic constraints commercial news outlets face. It is the perfect place for experimentation. All the Master of Science students already have a required data journalism class. That could also be incorporated into the work flow of the Media Center so that the stories written for the class would actually be published on Annenberg Media.

In order to broaden the scope and impact of data-driven journalism in Annenberg, future research topics could include holding local public officials accountable by data-driven means. In addition to topics generated by Crosstown Traffic, it would be possible to use open data provided by the Los Angeles County and its cities accompanied with California Public Record Act requests.

USC Viterbi might also be able to help in trying to make the Media Center more efficient as an organization and work place. They could come up with ideas, such as programs or applications, that might solve problems the Annenberg faculty and students are struggling with. The faculty and students of both institutions should participate in trying to come up with possible ideas for collaboration. One option would be trying to find solutions to problems the whole media industry is struggling with. It is an ambitious goal but why not have a go.

Both USC Annenberg and USC Viterbi are leaders in their own field. Together, they can pave a much more robust path to the future. The same holds true to the collaborative efforts of journalists and engineers in newsrooms all around the world.

BIBLIOGRAPHY

A Day in the Life of Uber. 2014. A data visualization by the Uber Data Team. Accessed May 11, 2017. https://newsroom.uber.com/a-day-in-the-life-of-uber/

Ali, Tanveer. 2014. "Local news outlets are joining the data journalism bandwagon." *Columbia Journalism Review,* August 7. Accessed May 9, 2017. http://archives.cjr.org/data_points/regional_data_journalism_becom.php

Bartlett, Rachel. 2013. "10 tips for compelling digital storytelling." *Journalism.co.uk,* May 22. Accessed 10 May, 2017. https://www.journalism.co.uk/news/10-tips-for-compelling-digital-storytelling/s2/a552987/

Bell, Emily. 2016. "The Tech/editorial culture clash." *Columbia Journalism Review,* fall/winter. Accessed 10 May, 2017. https://www.cjr.org/analysis/tech_editorial_facebook.php

Bradshaw, Paul. 2012. "What Is Data Journalism?" In *Data Journalism Handbook,* edited by Jonathan Gray, Liliana Bounegru and Lucy Chambers. Online version 1.0 Beta. Accessed May 10, 2017. http://datajournalismhandbook.org/1.0/en/introduction_0.html

Cherubini, Federica. 2015. "Trends in Newsrooms: Analytics, audience development and the newsroom." *World News Publishing Focus,* July 21. Accessed May 8, 2017. https://blog.wan-ifra.org/2015/07/21/trends-in-newsrooms-analytics-audience-development-and-the-newsroom

Codrea-Rado, Anna. 2012. "What Can Journalism Learn from Computer Science?" *Tow Center for Digital Journalism,* August 30. Accessed April 29, 2017. http://towcenter.org/what-can-journalism-learn-from-computer-science/

Gilmore, James H. and B. Joseph Pine II. 1997. “The Four Faces of Mass Customization.” *Harvard Business Review*, January-February. Accessed May 11, 2017. https://hbr.org/1997/01/the-four-faces-of-mass-customization

Government Salaries Explorer. *The Texas Tribune.* Accessed May 10, 2017. https://salaries.texastribune.org/

Leonhardt, David, Jodi Rudoren, Jon Galinsky, Karron Skog, Marc Lacey, Tom Giratikanon and Tyson Evans. 2017. ”Journalism That Stands Apart.” A report of the 2020 Group at *The New York Times*, January. https://www.nytimes.com/projects/2020-report/

Mapping L.A. *The Los Angeles Times.* Accessed May 11, 2017. http://maps.latimes.com/neighborhoods/

Matsa, Katerina Eva and Kristine Lu. 2016. ”10 Facts about the Changing Digital News Landscape.” *Pew Research Center*, September 14. Accessed May 9, 2017. http://www.pewresearch.org/fact-tank/2016/09/14/facts-about-the-changing-digital-news-landscape/

Mitchell, LaCrai. 2015. “How to Write an Unforgettable Data Journalism Story.” *IRE Conference Blog*, June 18. Accessed May 9, 2017. http://www.ire.org/blog/ire-conference-blog/2015/06/18/how-write-unforgettable-data-story/

Nelson, Laura J. 2016. “A tax hike to fund a major expansion of the Metro system is leading in early returns.” *The Los Angeles Times*, November 9. Accessed May 9, 2017. http://www.latimes.com/local/lanow/la-me-ln-metro-sales-tax-20161108-story.html

Pilhofer, Aron. 2012. “To Visualize or Not to Visualize?” In *Data Journalism Handbook*, edited by Jonathan Gray, Liliana Bounegru and Lucy Chambers. Online version 1.0 Beta. Accessed May 10, 2017. http://datajournalismhandbook.org/1.0/en/delivering_data_0.html

Pilsen Traffic Volumes. A data visualization by Open Transport Net. Accessed May 11, 2017. http://opentransportnet.eu/web/guest/pilsen-traffic-volumes

Quick Facts, City of Santa Monica, California. *United States Census Bureau*. Accessed November 5, 2016. https://www.census.gov/quickfacts/table/INC110215/0670000

Quick Facts, East Los Angeles CDP, California. *United States Census Bureau*. Accessed November 5, 2016. https://www.census.gov/quickfacts/table/INC110215/0620802

Schleuss, Jon and Ben Poston. 2016. “Los Angeles street quality grades.” *The Los Angeles Times*, August 14. Accessed May 10, 2017. http://www.latimes.com/projects/la-me-street-quality-grades-082016/#nt=outfit

Schmitz Weiss, Amy, Cindy Royal. 2013. “At the intersection of journalism, data science, and digital media: How can j-schools prep students for the world they’re headed into?” *Nieman Lab*, July 26. Accessed April 26, 2017. http://www.niemanlab.org/2013/07/at-the-intersection-of-journalism-data-science-and-digital-media-how-can-j-schools-prep-students-for-the-world-theyre-headed-into/

Sutedjo-The, Raymon and Sandra Lee. 2013. “Transit Quality and Equality: San Francisco.” A competition entry for the Urban Data Challenge. Accessed May 11, 2017. http://ray-mon.com/urbandatachallenge/sanfrancisco.html

Taylor, Brian D. 2016. An interview by Jutta Hoegmander. A phone interview October 31, 2016.

Thompson, Ben. 2015. ”Popping the Publishing Bubble.” *Stratechery blog*, September 16. Accessed May 8. https://stratechery.com/2015/popping-the-publishing-bubble/

APPENDIX A: TRANSCRIPTS OF THE STORIES

A.1 SCHEDULED DELAYS

The chronic congestion in Los Angeles is poisonous for buses. Metro's most delayed bus routes in the Los Angeles County all suffer from the same problem: traffic.

Three of the top five lines running most behind the schedule shuttle between Downtown and Santa Monica. Of those the 720 continues all the way to East Los Angeles.

"The Westside traffic problems are our biggest concern," says Robert Holland, the Sr. Executive Officer of Bus Operations at Metro Los Angeles, adding, "You can't get around in Santa Monica." (Holland 2016)

Holland says that some of the longest lines, especially the 720, which runs for 20 miles between Santa Monica and East Los Angeles, could be cut to half next summer (2016). That would improve their on-time performance and also increase the safety of the bus drivers. The drivers of the 720, for example, would not have to drive for two hours before having a breather.

The other two of the five worst bus routes in terms of delays are the 105 that runs from South Central along Vernon and up La Cienega to West Hollywood, and the 794 that begins its journey from Downtown and travels north to the San Fernando Valley.

"We have to fight all the traffic. It's tough," Holland says (2016).

Downtown has its own set of problems. There is a lot of traffic during rush hour and construction work, such as building the new regional connector, further worsens the situation.

The best performer of all the Metro buses is the Orange line that runs according to schedule 92 percent of time. The reason is clear: it has its own dedicated lane, like trains.

“That just tells you, if you keep the cars away from the buses we could run. But it's awful tough to do so,” Holland says (2016).

To separate the buses from the congested car traffic, the transportation authority has implemented designated bus lanes where possible. Holland says they work as long as they are clear. Unfortunately, the lanes are often clogged up by passenger cars. Holland says Metro tries to get the Los Angeles County Sheriffs Department, which is in charge of bus security, to be more aggressive in ticketing motorists who interfere with bus lanes. (2016)

According to the new Quality of Life report by Metro, the on-time rates of its buses have increased 13 percent since 2008 (2017). During the past couple of years, the rates have, however, been slowly declining. The official target is for 78 percent of the buses to run on time. Based on Metro’s own statistics, the on-time performance was nearly 76 percent in 2013 and a year later 74.5 percent. In 2015, the most recent full year for which statistics are available, 73 percent of the buses reached the goal. (Holland 2016) The reality may be even worse: According to Metro, a bus is considered on time if it leaves a stop a minute early or up to five minutes late in relation to the schedule.

A.2 The Unbearable Lateness of Buses

Arleen Ayala is waiting for one of Los Angeles’s most delayed buses at one of the most delayed stops: the Commerce stop in East Los Angeles of the 720 line.

What awaits her on the other end of the line is no better. She is going to ride the 720 all the way from East Los Angeles to its final stop in Santa Monica, just half a mile from the beach.

What these two places have in common is that they are among the worst spots in all of Los Angeles County for bus delays. An analysis conducted by the USC Viterbi School of Engineering's Integrated Media Systems Center and the USC Annenberg School of Journalism shows that the top two stops for bus delays are both in East Los Angeles. Three of the top ten stops where buses run most behind the schedule are in Santa Monica.

Ayala's 20-mile trip through Downtown Los Angeles is going to take two hours. That is, if the bus runs on schedule.

"It's a long time, but what can you do," Ayala says and shrugs her shoulders (2016).

The former East Los Angeles resident has been visiting her friends and family and is now on her way to meet her husband in Santa Monica, where he works. The family has only one car, so Ayala has to hop on the bus. They will continue the trip together back to their home near Bakersfield.

In a county that is notorious for its chronic congestion and dependence on cars, around one million people ride Metro buses on an average weekday. The 720 is one of the longest and most crowded lines.

But delays impact residents in East Los Angeles much more for the simple reason that they ride the buses much more than people who reside in Santa Monica.

According to the 2010 Census, 97 percent of the people living in East Los Angeles, a swath of the city nestled between Boyle Heights and Montebello, are of Hispanic or Latino origin. The area's median household income is $38,000, with a quarter of the population living below the federal poverty line. In comparison, people in Santa Monica have a median household income twice that of East Los Angeles, and three out of four are white.

In Los Angeles, as in other parts of the country, public transit is primarily the domain of the poor. Los Angeles Metro's bus customer survey shows that the median household income of the respondents was just below $15,000. In addition, over 80 percent of survey participants did not have access to a car.

Ayala says that buses are usually packed in East Los Angeles because public transportation there is crucial for people. "The rent is high in East Los Angeles. It's hard to afford a car or multiple cars and a house at the same time" (2016).

Prof. Brian D. Taylor, the director of the University of California, Los Angeles, Luskin School's Institute of Transportation Studies, says most riders are at the mercy of the bus system, even if it has chronic delays, because they simply don't have other options (2016). Public transit is in the US an inferior good. "It means that in general, as your income goes down, you consume more of it," Taylor says (2016).

Bus driver Albert Ballin knows the route of 720 by heart. He used to live in East Los Angeles and has been driving the line for six months now. One shift takes 11 hours, which means he drives the 20 miles route back and forth twice.

"There is a lot of construction and detours on this route. And traffic. We are now running 20 minutes late. That's quite normal on a Friday morning," Ballin says (2016).

Metro's bus schedules are revised every six months. If the on-time performance of the route is low, the schedule might be adjusted to make it more realistic. That can mean that a bus is technically "on time" even if it is crawling through traffic. Still, the 720 struggles to meet its mark on a regular basis.

According to Metro's own statistics, the 720 is on time nearly 70 percent of the time (Holland 2016). For Los Angeles Metro, buses are considered on time when they depart from the bus stop one minute early or five minutes late in relation to the schedule.

Ballin says that passengers get upset by the frequent delays, but there's not much he can do about it. "The traffic in Los Angeles doesn't stop. There's just too many cars." (2016) The 720 is a Metro Rapid bus, which is supposed to guarantee that patrons get to their destination more quickly than locals. Rapid lines have fewer stops, more frequent service and a transit signal system that is designed to reduce the time buses spend waiting at red lights.

Buses that mix with road traffic are most vulnerable to delays. Thus, the Metropolitan Transportation Authority has tried to improve the reliability and travel times of the rapid buses with dedicated lanes during rush hour.

There are people who rely on public transportation in Santa Monica too. One in 10 Angelenos who are employed say their main means of getting to work is public transportation. Bryan Plumer travels from Santa Monica to Downtown daily to his job in phone sales. He does not have a car because he trying to save.

Today Plumer had to wait for the bus for nearly half an hour before it showed up, 20 minutes late.

"As I was waiting, I saw four or five 720s pick people up on the other side of the street," Plumer says (2016). He notes that in the mornings, the service seems better westbound to Santa Monica than eastbound to downtown.

In fact, between 6 am and 7 pm there are a dozen buses leaving from East Los Angeles compared with four that start their journey from Santa Monica.

There is a reason for that. When the bus leaves from Santa Monica in the morning, it is half empty. More and more people hop on the bus as downtown approaches. After Wilshire and Vermont, there is an influx of Spanish-speaking patrons who fill the bus with a pleasant chatter. Soon, there are no seats left.

There is no display that indicates when the next bus will arrive at the most delayed bus stop in the whole of Los Angeles County, the Whittier–Hoefner stop in East Los Angeles. Emily Renteria, who is heading home with her nephew, uses one of many mobile apps that tell when the bus is supposed to arrive.

“Usually it’s not correct,” Renteria says (2016).

That does not matter because the buses run frequently enough to satisfy her needs.

Taylor, the UCLA professor, says that knowing when the next bus arrives is even more important to riders than running on schedule. Thus, the emergence of next bus indicators and apps that estimate the next arrivals. (2016)

Apel Bravo lives in East Los Angeles and rides the 720 bus daily to work at a seafood restaurant in Koreatown. Bravo knows that it is hit-or-miss whether the bus is delayed or not, so he leaves his house early to make it to work on time. If he needs to be at work at 4 pm, he has to leave at 2 pm. The 10-mile trip can take from 45 minutes to an hour and fifteen minutes depending on traffic. Despite the delays, Bravo isn’t complaining: “At least they’re working on it. They’re building more subway rails, adding lines, changing the schedule. But what can you do about traffic?” (2016)

A.3 REFERENCES

Ayala, Arleen. 2016. An interview by Jutta Hoegmander, October 4, 2016.

Ballin, Albert. 2016. An interview by Jutta Hoegmander, October 4, 2016.

Bravo, Apel. 2016. An interview by Jutta Hoegmander, October 4, 2016.

Department of Motor Estimated Vehicles Registered by County for the Period of January 1 through December 31, 2015. DMV Forecasting Unit. https://www.dmv.ca.gov/portal/wcm/connect/add5eb07-c676-40b4-98b5-8011b059260a/est_fees_pd_by_county.pdf?MOD=AJPERES

Holland, Bob. 2016. An interview by Jutta Hoegmander. A phone interview November 29, 2016.

How is Metro measuring up? 2008 – 2015 Quality of Life Report: The Full Report. LA County Metropolitan Transportation Authority. https://media.metro.net/docs/report_qualityoflife.pdf

Interactive Estimated Ridership Stats. LA County Metropolitan Transportation Authority. http://isotp.metro.net/MetroRidership/Index.aspx

Plumer, Bryan. 2016. An interview by Jutta Hoegmander, October 4, 2016.

Quick Facts, City of Santa Monica, California. *United States Census Bureau*. Accessed November 5, 2016. https://www.census.gov/quickfacts/table/INC110215/0670000

Quick Facts, East Los Angeles CDP, California. *United States Census Bureau*. Accessed November 5, 2016. https://www.census.gov/quickfacts/table/INC110215/0620802

Renteria, Emily. 2016. An interview by Jutta Hoegmander, October 4, 2016.

Spring 2015: Metro BUS Customer Satisfaction Survey Results (May 21-June 16, 2015). LA County Metropolitan Transportation Authority. http://media.metro.net/projects_studies/research/images/annual_survey_results/results_bus_spring_2015.pdf

Taylor, Brian D. 2016. An interview by Jutta Hoegmander. A phone interview October 31, 2016.