UNIVERSITY of WASHINGTON

NTIA Digital Nation Analysis and Visualization

Final Report

That Better Team (#19)

Janani Kumar, Nan Xiao, Shiyi Luo

Contents

EXECUTIVE SUMMARY

We believe universal information access is the first step to an egalitarian society**.** The easiest way to disseminate information is through the Internet. The National Telecommunications and Information Administration, an agency of the US Department of Commerce works towards providing Internet for everyone in America. The team worked closely with policy makers and analysts at National Telecommunications and Information Administration and the State of Washington: Office of Chief Information Officer to expose insights from the Digital Nation dataset, helping them arrive at data backed solutions to provide information access for all. Digital Nation data is collated by Bureau of Census by surveying 56,000 households across America on their computer and internet usage. The results of the data analysis will accelerate the goal of expanding Broadband access and adoption in America while ensuring internet remains an engine for continued innovation and economic growth.

SPONSOR ORGANIZATION BACKGROUND

NTIA:

National Telecommunications and Information Administration(NTIA) is an agency whose goal is to understand telecommunication and information infrastructure across the United States and accordingly advise the President of the United States on information policies. NTIA is not a stand alone agency but is a part of the US department of Commerce. The goal of NTIA is democratizing information access. As information access has the potential to change lives through innovation and education and is vital to progress as a nation, NTIA is devoted to studying internet usage , promoting broadband and improving access to internet.

State of Washington, Office of the Chief Information Officer:

The vision of OCIO is to be known as:

- Wise stewards of the taxpayer dollars
- Bold enough to make a difference
- Adaptive because we embrace and leverage a changing world
- Smart risk takers who experiment wisely with new technology and services
- Transparent and inclusive in our operational and financial priorities
- Technology leaders whom people trust because we exceed expectations
- Advocates who promote the successes of our state IT community

INFORMATION PROBLEM

Given that the goal of NTIA is ensure access to information through internet, they have been collectiNG information to understand current state of affairs every year. Working in conjunction with the U.S. Census Bureau's the NTIA conducts in person surveys and information collection campaigns through phones to collect information about internet usage in close to 50,000 households across the US. In the words of the project sponsor 'We touch every household at least 8 times to collect accurate information'. Also the amount of information collected per household is large. Given the massive data collection effort, there is an equally important effort required to unwrinkle and understand the data. At the moment, NTIA's efforts to visualize the data has resulted in dashboards that mask the richness in the data. NTIA also encounters difficulties in collecting accurate information from the remote regions of the country. This is information collection from remote regions is crucial to correct and introduce information policies that will benefit everyone.

There is a tremendous dataset that has the potential to ensure data driven information policies that can make information access reliable, easy and inexpensive. **Our goal is to build methodologies that will allow our stakeholders,who hold the positions of information policy analysts, to uncover important insights and share it with decision makers in the government. We will eliminate their current difficulties and create interactive visualizations that will allow the data consumer to engage powerfully with the dataset. Sub tasks in this main task involve finding data stories and publishing blogs on the same.**

RELATED EXISTING SOLUTION

1.**Digital Nation Data Explorer**: This is a data visualization web application that enables easy tracking of metrics about computer and Internet use over time produced by NTIA. We can choose a metric of interest from the drop-down menu. The default Map mode depicts percentages by state, while Chart mode allows metrics to be broken down by demographics and viewed as either percentages of the population or estimated numbers of people or households. Meanwhile, the weakness of the project is that it lists out large volume of data in the dataframe below the D3 visualization, which is not very intuitive and readable.

2.**Census Data Visualization Gallery:** This is a visualization gallery that the Census Bureau worked on to make public data available for any user. The first posted visualizations will pertain largely to historical population data, building on prior work done to portray historical changes in the growth and redistribution of the U.S. population. For later visualizations, the topics will expand beyond decennial census data to include the full breadth of Census Bureau data sets and subject areas, from household and family dynamics, to migration and geographic mobility, to economic indicators.

3.**Broadband Adoption in WA:** This is a Tableau visualization that our sponsors created for broadband usage of Washington State density. It uses bar charts to show the broadband adoption of different devices and location from 1998 to 2015, as well as comparing broadband adoption statewise among WA, OR, ID and MA. While the bar size and color hue indicate the adoption for each user group clearly, we think more visual variables can be added to make the work more informative and interactive. Also, we aim to add county level analysis and visualization based on that.

OUR DATA SOLUTIONS

To address the problem states before, three tasks were created to solve it.

Task 1:

State Broadband Profiles: Create profiles for computer and internet use in each state, and present the information using compelling visuals and infographics that can inform policy at the state level. This is to help state level agencies to understand internet usage readily and allow comparison with other states. They will have all information at once without having to run queries or wait on accumulating data.

Task 2:

Find a New Story: Uncover interesting patterns about Computer and Internet use in America from the Data. The data has about 1000 attributes. The goal is to expose and analyze important attributes that will aid better policy making

Task 3:

Dynamic Visualizations: To understand existing digital divide, dynamic visualizations that allow policy makers to explore the data had to be built

And Make it all Repeatable: Develop necessary tools, code, and documentation so that people can continue the work after project is complete.

Digital Nation Data 2015

The dataset is huge:

```
entire_dataset=pd.read_csv("C:/Users/jakum/Documents/Capstone/data/data_july15.csv")
```

```
#total number of attributes in the data 978
#480 of integer type and 498 of float type.
dtp=entire_dataset.dtypes
dtp.value_counts()
```

```
float64    498
int64      480
dtype: int64
```

```
#73395 unique households across US were sampled for the survey
print(entire_dataset.qstnum.nunique())#unique house id
print(entire_dataset.occurnum.nunique())#unique householder id
```

```
73395
15
```

```
#to identify each record in the dataset uniquely, the variables qatnum and occurnum can be used
temp=entire_dataset["qstnum"].map(str)+entire_dataset["occurnum"].map(str)
print(temp.nunique())
```

```
149414
```

Validating Our Data Processes

To ensure validity of our computation, we computed the estimates on the Digital Nation Data Explorer and they matched. We had the guidance of Rafi Goldberg, who owns the data set and is a policy analyst at NTIA to help us make correct estimates.

Task 1: STATE INFOGRAPHICS

We created state level one pager infographics to show computer and internet user profile. Information covered are user profile divided by internet devices, locations, purposes, user demographic profiles and children internet usages.

Starting from the top left of the infographic, we can first see a bar chart representing the percentages of users using different internet devices. Beside is another bar chart showing the reasons that people don't want to use the internet. Below that is series of bubbles describing the digital divides of specific groups of people, including their internet usage information and a comparison with national level.

On the right part of the infographic is the demographic profile of households with access to computer and internet. Here we can see the gender, age, household size, annual income level and education level of people. We observed a significant difference between the average household size and income level of households with and without internet access. Households with internet access have larger size, higher annual income and education level.

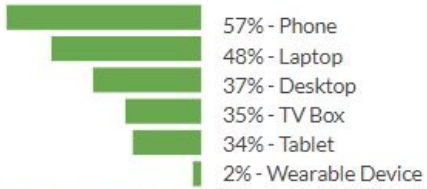
At the bottom is the profile of children using internet. We can see most popular devices and locations children use for internet, how preschool age children and school age children use internet differently, the state children internet habits comparing to national level, and their reasons of not using the internet.

Digital Nation

Washington State Profile

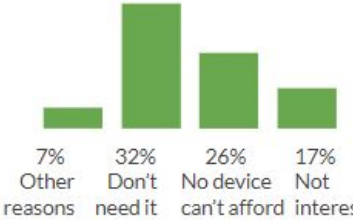
(Data from Digital Nation 2015)

Devices people use to get on internet:



Why

don't you want to use computer and internet at home?



Digital Divides

In the Nation and Washington State: ***593,679 households are not using the internet***

- **19%** Of adults (22+) don't use internet, and 25% don't use it at home.
- **31%** Of people 15 or over with HS degree or less don't use internet.
- **31%** Of people 15 or over with a disability don't use internet.
- **36%** Of people over 65 or older don't use internet.

Demographics

Profile of Internet Households

Gender

49.8% 50.2%

Household Size

Avg. 3.3

Income and Education Level

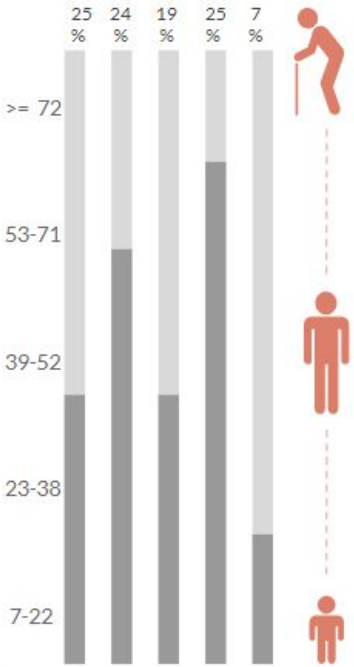
Avg. $75,000

Households with access to internet and computer have annual median HH income of 75k.

(Average Overall is 63k)

- *25% hold a High School Degree*
- *33% hold a Bachelor's or higher degree*

Age



How do the **Children** (3 - 12 years old) use internet ?

77.5% of children have access to internet

Most used mobile device:

Phone

77% use phone
55% use laptop
3% use wearable devices

Location of using Internet:

		Pre-schoolers	School age
Home	WA	46.4%	67.8%
	US	40.7%	60%
School	WA	8.3%	49.3%
	US	17%	49%

- *Older children use internet more often at school.*

Major reasons children don't use internet:

1. Don't need it
2. Not interested

Infographic for State of Washington

Making Infographics Repeatable:

We designed and created the Washington State Computer and Internet User Profile as the sample infographic. The same template can be used to create state level profiles for other state and other updated data source. To reduce the manual process and automize the reproducing work, we also created a Data Exploration Tool using Tableau to help us get the key metric numbers for each state.

The Data Exploration Tool contains the digital nation dataset 2015 and all data aggregation and pre-processing steps. It has two dashboards, the first one contains the computer and internet usage numbers, and the second one contains the demographic profile numbers. Following the instruction on the dashboard, one can filter the data by state and get the numbers needed for infographic for any state. To reproduce an infographic for another state, one needs to open the Data Exploration Tool first, filter by that state, get the updated numbers, then follow the instructions to do some simple calculation in order to insert the final numbers into the infographic template.

Task 2: DATA STORIES

The goal of data stories is to expose parts of the data that uncover meaningful insights that have not been explored before. For this work, focus was upon exploring attitude towards security, incidents of cyber crime and cyber bullying and digital divide. An attempt to answer the following questions was made:

1. What are the main reasons Americans state for not using the Internet?
2. When they do use, how safe do they feel on the Internet?
3. What fears do they have about using Internet for certain activities?
4. What are the basis of their fears?
5. What are they scared of doing on the Internet?

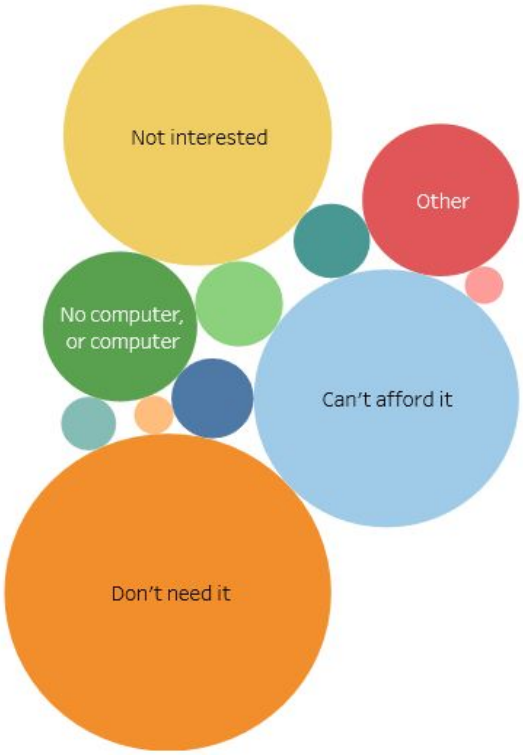
Data wrangling for the data stories was performed using Python and Tableau was using to visualize the data stories.Next three pages will have the resulting visualization of data stories.

Reasons for not using Internet Across US

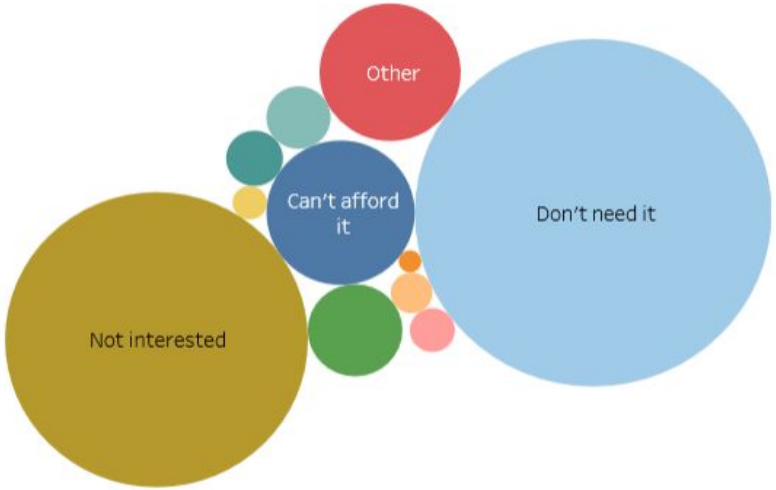
Among households that do not have internet access at home, a variety of reasons were given. Here are the main reasons provided by the "reference person" in affected households, who are considered to be the head of the household. The average age of these reference persons are is 57.7.

How to interpret the data: Hover over each bubble to learn more. As an example, 22.27% of household that do now have internet at home said they were not interested in using Internet;25% of households that do not have internet at home said they were not interested in using internet outside home.

At Home



Outside Home



Privacy and security concerns among Internet users across US.

This dashboard highlights the privacy and security concerns expressed by internet users across US and by state.
Click on a concern to highlight the corresponding numbers in the vizualizations. Use the state filters to learn more about a particular state.

Concern
- Others
- Threats to personal safety such as bullying
- No concerns
- Data collection or tracking by government
- Loss of control over personal data such as e...
- Data collection or tracking by online services
- Credit card or banking fraud
- Identity theft

State
- (All)
- AK
- AL
- AR
- AZ
- CA
- CO
- CT
- DC
- DE

Across US

By State

Online Activities hindered due to Privacy and Security Concerns.

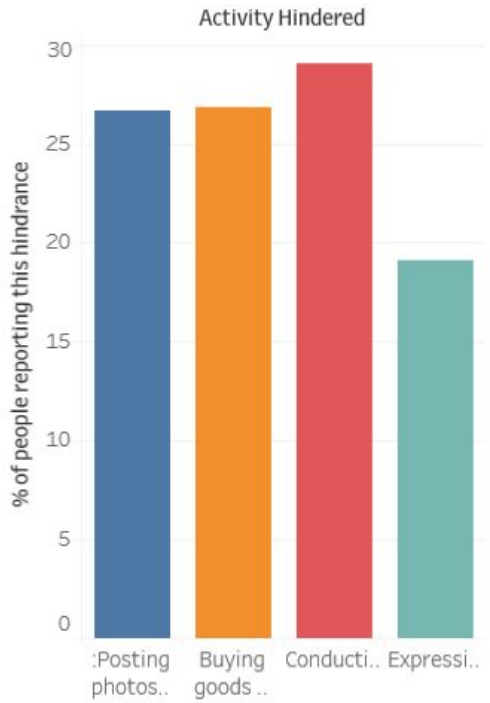
Americans use of Internet for certain activities has been hindered by privacy concerns.
How to interpret the data: Each color represents one such activity hindered by privacy and the size of the bubble represents the % of internet using households that expressed the same. Click on an activity to highlight its numbers. Use state filters to learn more about a particular state.

Example : Across US 29% of internet using housholds don't conduct financial transactions due to fear of cyber crimes; 27.6 % of all internet users in WA, 20% of all internet users in ID.

Activity Hindered
- :Posting photos, status updates, or otherinformation on social netwo...
- Buying goods or services online?
- Conducting financial transactions such as banking, investing, or payin...
- Expressing an opinion on a controversial or political issue on a blog or...

State
- (All)
- AK
- AL
- AR
- AZ
- CA
- CO
- CT
- DC
- DE
- FL
- GA
- HI
- IA
- ☑ ID
- IL
- IN
- KS
- KY
- LA
- MA
- MD
- ME
- MI
- MN
- MO
- MS
- MT
- NC
- ND

Across US



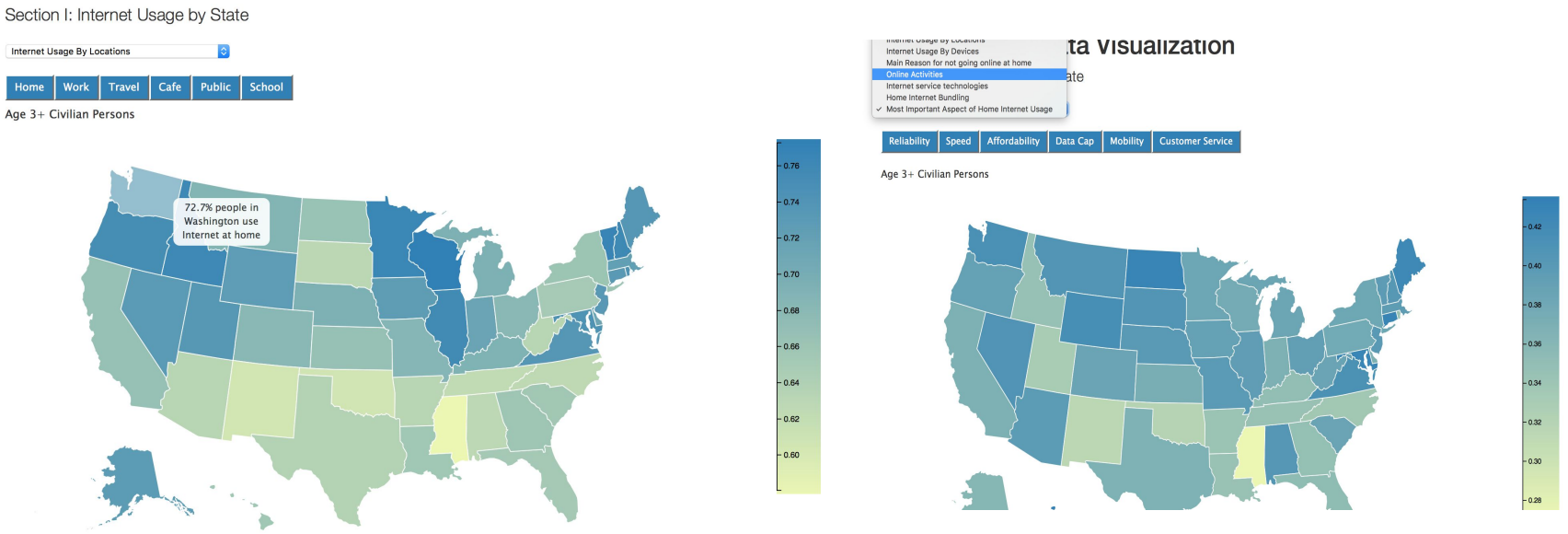
By State

Task 3: DYNAMIC VISUALIZATIONS

The dynamic D3 visualizations contain explorations to understand digital divide.It contains two sections.

Section I

The first section is a choropleth map that enables users to capture difference and compare Internet usage information across different states in the US. The visualization gives general overview of main attributes in the digital nation dataset.



VISUALIZATIONS DESIGN AND FEATURES FOR SECTION I

a.Color code: Blue represents high value and yellow represents low value. As shown in the demo graph above, the states in blue means the Internet usage percentage is above national average level; on the contrast, yellow means the Internet usage percentage is above average level. Greenish color represents the value for that state is close to national average level.

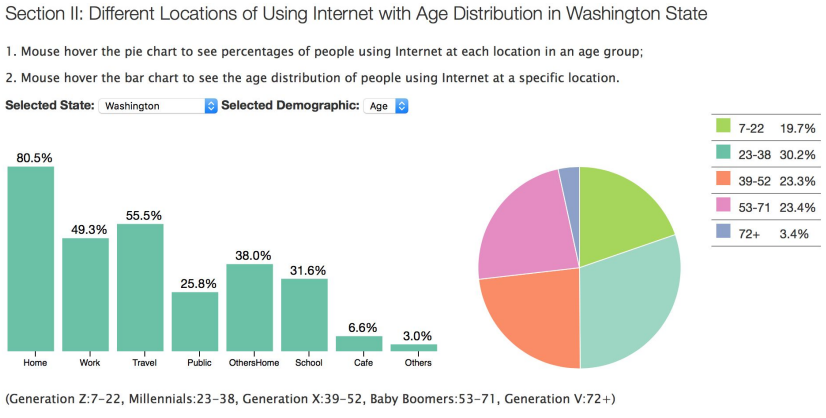
b. Dropdown selector: From the dropdown on the top left, users can select from seven topics they want to explore. The topics include “Internet usage by location”, “Internet usage by devices”, “online activities”, and “main reason for not going online at home”, etc. When one topic is selected from the dropdown selector, the buttons will change accordingly showing all options to the specific question.

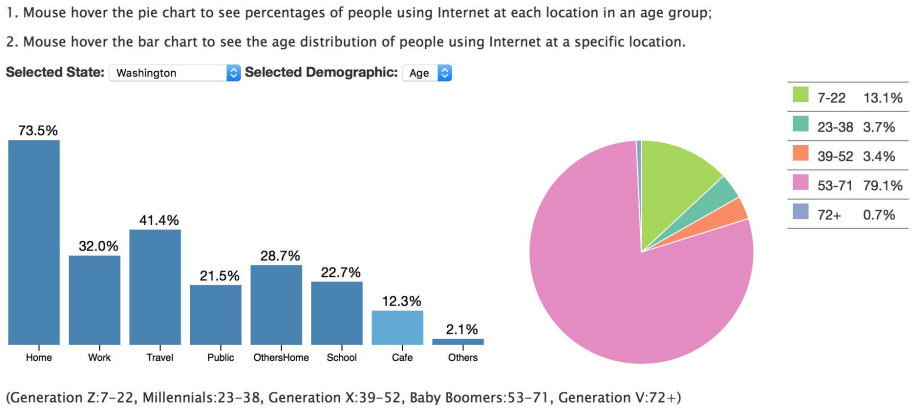
c. Buttons: Each button present one measure and option, clicking on the button will change the choropleth map accordingly. For example, when click on “home” button, the visualization will display the percentages of people use Internet at home in each state.

d. Tooltip: Hover on state to show the exact percentage value, the labels with Internet usage percentage data appear when hovering on state and disappear when mouse out.

Section II

In section two, we added one more dimension with demographic information into the visualizations. In this section, users could explore the correlation between any topic
(Internet usage by location) and demographic information (age, sex). The bar chart and pie chart can be mutually filtered with the other visualization.





VISUALIZATIONS DESIGN AND FEATURES FOR SECTION II

Visualizations design and features:

a.Color code: Each color represents one demographic group. The color is applied to both bar chart and pie chart.

b. Dropdown selector: Users can select specific state and demographic information they are interested in by selecting them from the two dropdowns. There are 50 states and two demographic options available.

c. As shown in the graph, when hovering over on one slice on the pie chart, the bar chart on the left will display the Internet usage information by location for the selected age group. For example, top 2 places are home and traveling: 80.5% of people aged from 23-38 used Internet at home and 55.5% of them used Internet while traveling between places.

d. From another perspective, when hovering over on one bar on the bar chart, the pie chart on the right will display the age group percentage distribution for people used Internet at selected location. For example, Among people used Internet in Cafe, 79.1% of respondents are aged from 53-71 and 13.1% of them are 7-22 years old.

TRANSITIONING PROJECT

Overview of next steps/your plan to transition your project output to the sponsor/organization

1. Hand over our code, visualization template, data processing tools,technical documentation by June 8th.

2. Link to all of our deliverables

Code repository:

https://github.com/nannxiao/ntia

D3 Visualization Link:

https://nannxiao.github.io/ntia

Tableau Data Story Dashboard:

https://public.tableau.com/profile/janani.kumar#!/vizhome/DataStories/Story1?publish=yes

State Level User Profile:

https://drive.google.com/open?id=1v8sYILuGAf5dtgZRTSC059Y82tPIOFHz-A4upuCmUKg

State Level User Profile Data Exploration Tool:

https://drive.google.com/a/uw.edu/file/d/1X1xRaiks8srFCrNB_lLZjFfVnfV_I_PK/view?usp=sharing

3. Our postgraduate contact information

Github:

@jkumar204

@nannxiao

@luos9

Linkedin:

https://www.linkedin.com/in/janani-kumar
https://www.linkedin.com/in/nan-xiao
https://www.linkedin.com/in/shiyi-luo-66494b93

IMPACT OF OUTCOME

Post hand-off the State and Federal Agency will absorb the work to delve deeper into the data. The latest digital nation data will be coming in the next few months and this work will be used as a framework to analyze the data. The team will be in touch with the sponsors to help them absorb this work over the next 6 months.

Citations

Digital Nation Datasets. Retrieved January 19, 2018, from https://www.ntia.doc.gov/page/download-digital-nation-datasets

Data Central. Retrieved January 19, 2018, from https://www.ntia.doc.gov/category/data-central

The Website Services & Coordination Staff, US Census Bureau. (1994, March 01). Data Visualization Gallery. Retrieved January 19, 2018, from https://www.census.gov/dataviz/

Office of Privacy and Data Protection. Broadband Adoption in WA. Retrieved January 19, 2018, from https://privacy.wa.gov/broadband-adoption-charts