Mathworks: Defeating the Digital Divide

Executive Summary

As a reaction to the COVID-19 pandemic, “Governments [have] issued 'stay at home' orders to reduce the spread of contagious diseases” [1]. This has caused increased reliance on the internet [2] for communication and has exposed inequalities in internet access [3] and its impacts on society in specific scopes of life: school, healthcare access, work from home, civic participation, and entertainment.

To determine the cost per unit of bandwidth in dollars per Mbps over the next years for consumers in the United States and United Kingdom, we used an exponential fit over 3 data points. We had data for median prices of connectivity and mean internet speeds for 2014 data, so we used a logarithmic regression to relate this information. This was viable because a t-test for difference in means revealed that the difference between the 2 data sets was not significant. Using a differential equation, we proved that the exponential decay model was our best regression fit, enabling us to devise a final equation. We used this equation to calculate the costs, ranging from $0.14 in 2021 to $0.01 in 2031. This is an optimistic metric for the future that can decrease disparity among internet connectivity for lower income and higher income residents.

Our next task was to approximate a household's need for internet over the course of a year, which was interpreted as Terrabits per year. First, we had values for hours per week of activities using internet connectivity based on age and income categories. Second, we used dimensional analysis for conversion, by multiplying vectors by Mbps for all the activities. Third, we constructed linear regressions for both factors individually. Since the linear regression using income had a significantly larger coefficient of determination this was the better model. Lastly, we used our model to demonstrate the minimum bandwidth in 3 situations 90% and 99% of the time.

Our last task was to develop a model that produced an optimal plan for distributing cellular nodes in a region. We developed a solution that attempted to minimize the cost of placing cell towers by varying the type of cell tower placed within each region. We then used the difference between the cost of implementing the tower compared to the value in price per Mbps it brought to the region over a one month period. We concluded

that the use of a low band tower would accomplish this task the best because it had wide coverage and was able to meet the predicted bandwidth needs of most households.

Global Assumptions

1. Terrabits/year is an accurate representation of internet needs.

2. Entertainment is an important part of bandwidth needs. 60% of all internet traffic is generated from streaming videos [4].

Global Definitions

Bandwidth: The maximum amount of data transmitted over an internet connection in a given amount of time [5].

Internet Need: We consider access to educational resources, healthcare, civic participation, and entertainment.

Part I: The Cost of Connectivity

Restatement of the Problem

We are tasked with creating a model that estimates the cost per unit of bandwidth in dollars or pounds per Mbps over the next 10 years for consumers in the United States and the United Kingdom.

Local Assumptions

1. The Mbps peak download speed is the unit that determines the cost of the bandwidth in dollars.
2. The rate at which the the price per Mbps of peak download speed follows an exponential decay model.
    - **Justification**: Based on the data from D2, the rate at which the cost of bandwidth in dollars changes is nonlinear and negative.
3. No outlier shock factors— such as widespread policy implementations or major changes in consumer behaviors —will produce non-negligible changes to the model in the future. Policy and regulatory, infrastructure, and supply and demand factors remain unchanged over the next 10 years.

- **Justification**: Outlier factors such as policy or regulatory changes, supply and consumer demand, and changes to infrastructure seem to affect broadband price [6][7]. These outliers are not easy to predict and some are qualitative thus making it hard to determine its effects.

4. The cost per unit of bandwidth between the United States and United Kingdom are similar enough to be combined into the same model, however the output of the model for the United Kingdom will be adjusted to be slightly higher.
   - **Justification**: There was not sufficient data to develop independent models for the United States and the United Kingdom. However, past research [8] demonstrates that the price per Mbps is lower in the United States than it is in the United Kingdom even though the United States generally pays more total.
5. Plans of increasing download speed will increase in price at a nonlinear rate that decreases as download speed increases.
   - **Justification**: Based on the data in D2, the price increases at each higher tier plan for download speed decrease, where download speeds grow disproportionately higher per dollar as the price of the plan increases.
6. The distributions of median monthly prices and average monthly prices were both normal.

Symbols Used

k - the rate of decay in price over time

t - time in years

A_0 - the price per Mbps of peak speed in the base year (2012)

$\hat{A}$ - the predicted value of price per Mbps of peak speed as it relates to time.

$\bar{p}$ - the average of the average monthly prices per peak Mbps of each city

b_d - the download speed in Mbps (bandwidth)

b_p - the peak download speed in Mbps (bandwidth)

s - standard deviation

n - number of elements in a series

μ - population mean of a distribution

Solution and Results

We ran an exponential fit over three data points. For years 2012 and 2020, we used the data from D1 to calculate the average monthly price per Mbps as such:

$$\bar{p} = \frac{1}{n}\sum_{i=1}^{n} p_i$$

We derived the point for 2014 using the data from D1, which gave us the average download speeds (10.5 and 9.9 for the US and UK respectively), allowing us to map them to the median montly prices given in D2, which were fit to a logarithmic regression as it provided the closest, continuously increasing fit to the data resulting in the following equation:

$$p_d = 14.5 + 11.4\ln(b_d)$$

This equation was used to estimate the price of average download speeds in the US and UK since they did not fall within the discrete set of ranges to median monthly prices found in D2. This resulted in predicted prices of 41.31 and 40.63 for the UK and US respectively. Using these prices, we then calculated the price per peak Mbps of download speed using the related values from D1, which were 40.6 and 42.2 for the US and the UK respectively. Thus the price per peak Mbps of download speed was $\$0.98$ and $\$0.97$ respectively, resulting in an average price of $\$0.98$.

Although we used the median monthly price in comparison to the average monthly price, we determined through statistical analysis that the difference in median and average prices were negligible.

We ran a 2 sample t-test for the 2020 data. We assumed that the distributions of median monthly prices and average monthly prices were both normal. Given 14 elements, we used the Central Limit Theorem (CLT) in order to justify that our sample sizes were large enough. We understood that the elements were independent because 14 times 10 is greater than or equal to 50. Thus, all of the conditions for the test were met.

Our null hypothesis (H_0) was that the mean for both populations of states did not vary. Our altenerative hypothesis (Ha) was that the mean for both populations of states differed.

$$H_0 : \mu_1 - \mu_2 = 0$$
$$H_a : \mu_1 - \mu_2 \neq 0$$

First we calculated the t-statistic. We knew $x_1$and x_2 were equal to .404 and .174 respectively.

$$t = \frac{(\bar{p}_1 - \bar{p}_2) - (\bar{\mu}_1 - \bar{\mu}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Our difference in population means is assumed to be zero. Using this formula, we obtained $t = 2.59$.

We calculated our p-value using the cumulative distribution function for our normal distribution. $P(\mu_1 - \mu_2 \neq 0) = 2P(X \geq 4.326) = 0.23$ with 13 degrees of freedom. With an alpha level of .01, then we do not have sufficient evidence to reject the null hypothesis.

Thus we were able to test it with both the average median monthly price per Mbps $\hat{p}_{med}$ and with the average mean monthly price per Mbps $\hat{p}_{avg}$ for 2012 and 2020 with the median average monthly price derived for 2014.

We considered an exponential decay regression since with dimensional analysis, we attempted to find the price as it relates to the change in peak Mbps. By rewriting the equation as follows:

$$\frac{\$}{\frac{Mb}{s}}$$
$$\frac{\$}{\frac{Mb}{s}} \times \frac{s}{s}$$
$$\frac{\$ \bullet s}{Mb}$$

We found that the the price can be directly related to the change in Mbps. Since the bandwidth increases faster than the cost, $db_p > d\bar{p}$, a negative constant k must exist such that:

$$\frac{d\bar{p}}{dt} = -kb_p$$

Thus we can derive an expontential relationship between the two with a decay rate k:

$$\int \frac{d\bar{p}}{b_p} = \int -kdt$$
$$\ln |b_p| = -kt + C$$
$$b_p = e^{-kt+C} = e^C e^{-kt}$$
$$b_p = A_0 e^{-kt}$$

Now we have evidence of exponential decay. This resulted in the following regression models:

$$\hat{p}_{med} = 2.15e^{-0.233t}$$

$$\bar{p}_{avg} = 1.67e^{-0.278t}$$

When evaluated, $\hat{p}_{med}$ resulted in an r^2 value of 0.906 and $\hat{p}_{avg}$ resulted in an r^2value of 0.999. The models consider 2012 to be the base year at $t = 0$. Thus the predicted cost per Mbps of peak data speed in the next ten years would be as follows:

Predicted Cost Per Mbps of Peak Data

Year	t value	Predicted Median	Predicted Mean
2021	9	$0.26	$0.14
2022	10	$0.21	$0.10
2023	11	$0.17	$0.08
2024	12	$0.13	$0.06
2025	13	$0.10	$0.05
2026	14	$0.08	$0.03
2027	15	$0.07	$0.03
2028	16	$0.05	$0.02
2029	17	$0.04	$0.01
2030	18	$0.03	$0.01
2031	19	$0.03	$0.01

Strengths and Weaknesses

It was very possible for the model to contain error from overfitting or with the predicted means since on point was a predicted median. The model was only fit with respect to three data points, which is generally a very small sample size to determine a continuous regression output. In addition, the model relies heavily on the assumption that the decay rate will not change significantly due to any unexpected outlier sources, and can be proven wrong with any sudden changes in consumer behavior or major non-negligible policy changes, regulations, or events. Furthermore, the use of average peak data instead of average download speed may have produced a less accurate model since peak speeds don't happen as often as average download speeds even though average peak speeds tend to rise as average download speeds rise.

The model also does not discern whether the drop in price is a result of a drop in overall price or the result of an increase in the amount of average peak Mbps in a particular city because it relies on the rate at which the price per Mbps varies with respect to time. In addition, the data and averages used to build the exponential model may not have been a representative model of the entirety of the US and UK given that they were only select cities with broadband utilities, and thus better infrastructure than places that tend to be more rural. This model will likely have a large source of error because it does not account for rural areas, but data in those regions are harder to collect because they lack the technological infrastructure available in larger cities, so it may be tough to even test the magnitude of that error in the future.

However, several parts of the methodology were supported by analysis: (1) the t-test supported the use of the median data point in conjunction with the averages in a model, (2) the choice of a non-linear exponential rate of change by derivation through integration, and (3) the high r^2 values as a result of the assumption of an exponential rate of change. Lastly the results seem to make physical sense with respect to the existing trends: (1) the median price has started to decrease slower than the mean price, and (2) both the mean and median price has been declining in a non linear fashion with a rate of change that slowly decreases. It would not be surprising that the Mbps available greatly increases and the price decreases as a result of future technological innovations and increased access, even to the point where 1 Mbps would only cost about a cent on average.

Part II: Bit by Bit

Restatement of the Problem

We are tasked with developing a mathematical model that will estimate a given household's internet need over a year and will then be applied to sample households.

Assumptions

1. The average weekly hour usage of data is representative of a whole year and the estimations for Mbps usage by category.
2. The following activities are negligible: frequent large file downloads and high quality video streaming.
   - **Justification**: Although the bandwidth used with frequent large file downloads is large, they take less time to download and are uncommon occurrences for the average person [9]. High quality video streaming is not considered necessary because standard definition videos suffice for standard educational and informational content.
3. Categories given by D5 that give required bandwidth amounts equate to the activities listed in D4.

   We equate:
   - "General web surfing, email, social media" to "Internet on a computer (not including video)"
   - "Online gaming" to "TV connected game console"
   - "Video conferencing" to "video on a computer"
   - "Standard definition video streaming" to "TV Connected Internet Device", "Total App/Web on a Smartphone", and "Total App/Web on a Tablet"
4. Our model does not account for the effects of the COVID-19 pandemic. This model instead measures the bandwidth usage in an average year with no major events, non-negligible outlier factors, or changes in consumer behavior.
   - **Justification**: The survey data used to construct the model (by estimating the Mb consumption) was not trained on data during the pandemic, but rather, before the pandemic so it cannot account for a pandemic event.
5. We used the average bandwidths to average income for each category and assumed that they were representative of total usage.

6. If the income and profession of a person is not stated, we opt for the use of average ages instead.
7. If the profession of a person is stated, then we use the average salary of that profession.
8. The average age of a single parent of a toddler is 26 + the age of the child.
    - **Justification**: The average of a single parent was computed by taking the average ages of having a first child for a woman and a man ([age first-time woman + age first-time man]/2) [10].
9. For some categories, data is omitted for the 2-11 and 12-17 age groups. We assume that the average hours per week is equal to data from the 18-34 age group.

Symbols Used

$\vec{A}_i$ - Vector whose ith term is the number of Mbps for the ith activity

- A_1 - The average bandwidth used for general web surfing, email, social media
- A_2 - The average bandwidth used for online gaming
- A_3 - The average bandwidth used for video conferencing
- A_4 - The average bandwidth used for standard definition video streaming

$\vec{B}$ - 2D vector whose ith term is a 1 by 6 vector where each entry is the hours per week spent for the ith activity for each jth age category

- The ith activities in B correspond to the ith activities in A
- $\vec{B}_1$ - Age-based vector used for Internet on a computer (not including video)
- $\vec{B}_2$ - Age-based vector used for TV connected game consoles
- $\vec{B}_3$ - Age-based vector used for video on a computer
- $\vec{B}_4$ - Age-based vector used for TV connected Internet devices, total app/web on a smartphone, total app/web on a tablet

$\vec{Y}$ - A 1×6 vector whose inputs represent the sum of hours spent on activities in hours/week for each age group.

$\vec{Y}_{avg}$ - A vector relating the average age to average bandwidth by activity

$\vec{Y}_f$ - The final vector relating the average age to average bandwidth by activity

$\vec{I}$- 2D vector whose ith term is a 1 by 4 vector where each entry is the hours per week spent for the ith activity for each jth income category

$\vec{X}$ - A 1×4 vector whose inputs represent the sum of hours spent on activities in hours/week for each income category.

$\vec{X}_f$ - The final vector relating the average income to average bandwidth by activity

H- The total Megabits being used by a household (over a year)

Solutions and Results

The required bandwidth per activity can be given by the following vector:

$$\vec{A}_n = \langle 1, 2, 2.5, 3.5 \rangle$$

We referred to the first table in D5 to compute these quantities by taking the average bandwidth (Mbps) for each activity.

First, we created a model based on age category. To do so, we considered the charts from D4 as the 2D vectors $\vec{B}_1$ and $\vec{B}_2$ representing the hours/week of each activity per age group in the 1st quarter of 2019 and 2020:

First Quarter of 2019

Aa Activity	2-11	12-17	18-34	35-49	50-64	65+
Internet on a Computer (not including video)	3.7	3.7	3.7	4.25	4.4	2.77
TV Connected Game Console	2.92	4.08	3.73	1.62	0.45	0.15
Video on a Computer	0.87	0.87	0.87	0.63	0.5	0.23
TV Connected Internet Device, Total App/Web on a Smartphone, Total App/Web on a Tablet	34.59	32.47	34.37	36	29.05	22.66

First Quarter of 2020

Aa Activity	2-11	12-17	18-34	35-49	50-64	65+

Activity	2-11	12-17	18-34	35-49	50-64	65+
Internet on a Computer (not including video)	3.97	3.97	3.97	4.77	5.03	3.17
TV Connected Game Console	2.72	4.18	3.63	1.73	0.47	0.17
Video on a Computer	1.73	1.73	1.73	1.28	1.07	0.5
TV Connected Internet Device, Total App/Web on a Smartphone, Total App/Web on a Tablet	42.27	39.07	41.5	44.12	37.45	30.29

The average bandwidths, A_1, A_2, A_3, A_4 are scalar quantities. We used the following equation to calculate the hours/week for each age group:

$$\vec{Y_1} = \langle \sum_{1}^{i=6} (\vec{B}_{i,j} \times A_i) \rangle$$

Using our data values, our new vector was:

$$\vec{Y_1} = \langle 132.78, 127.68, 133.63, 135.065, 108.24, 32.94 \rangle$$

The same process can be used to calculate the vector $\vec{Y_2}$:

$$\vec{Y_2} = \langle 161.68, 153.41, 160.82, 165.85, 139.71, 110.74 \rangle$$

Then we averaged the values from $\vec{Y_1}$ and $\vec{Y_2}$ to get the new vector $\vec{Y}_{avg}$:

$$\vec{Y}_{avg} = \langle (\frac{\vec{Y_1} + \vec{Y_2}}{2}) \rangle$$
$$\vec{Y}_{avg} = \langle 147.23, 140.55, 147.22, 150.46, 123.97, 96.84 \rangle$$

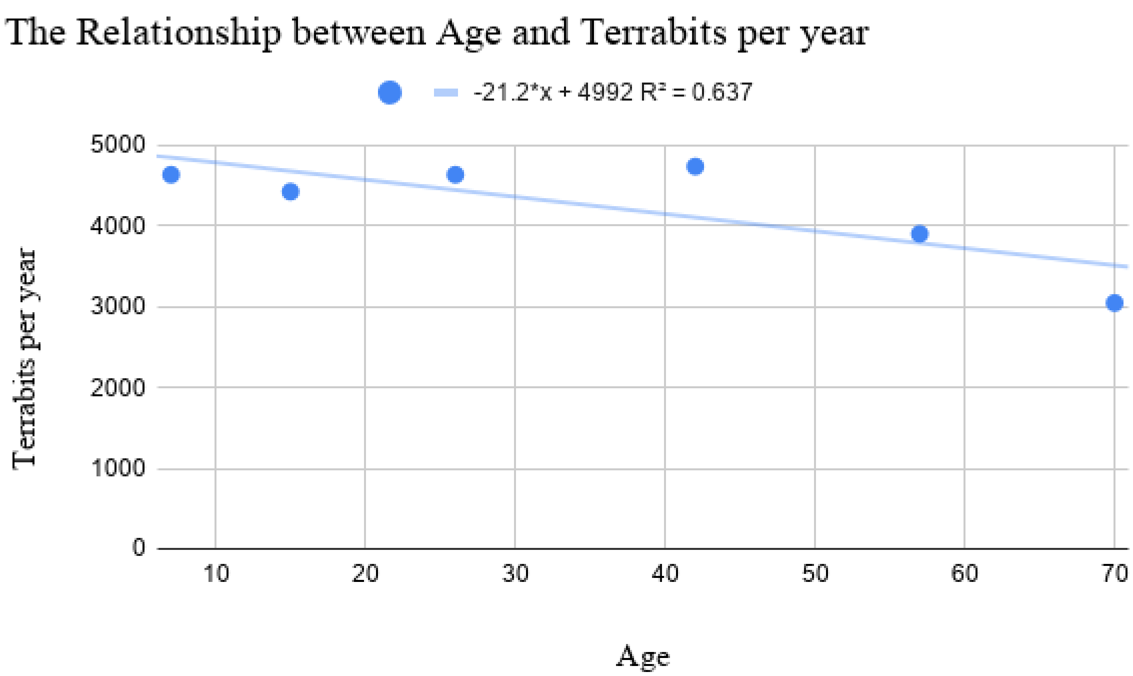
We used dimensional analysis to find the total Megabits per year:

$$B_{ij} \times \frac{52\ weeks}{year} \times \frac{3600s}{1\ hour} \times A_i = H$$
$$\frac{168\ hour}{week} \times \frac{52\ weeks}{year} \times \frac{3600s}{1\ hour} \times \frac{Mb}{s} \times \frac{Tb}{1000000Mb} = \frac{Tb}{year}$$

After unit conversion, we got the final vector $\vec{Y_f}$ in terrabits per year:

$$\vec{Y}_f = \langle 4630, 4420, 4630, 4731, 3898, 3045 \rangle$$

By relating this to the average ages per category, we constructed the following linear regression:



Although the R^2 value indicates that the line of best fit represents the data reasonably well, we can see by the scatterplot that there is more of a random dispersion. Therefore, we opted not to use a regression of age. Instead, we focused on constructing a regression based on income category.

We consider the remaining chart from D4 as the 2D vectors $\vec{I}$ representing the hours/week of each activity per income group in the 1st quarter of 2019 and 2020:

Third Quarter of 2015

Activity	< $25k	$25k-50k	$50k-75k	>$75k
Total Internet on a Computer	12.01	9.71	9.12	7.85

Activity	< $25k	$25k-50k	$50k-75k	>$75k
TV Connected Game Console	9.89	7.83	5.34	4.19
Total App/Web on a Smartphone	13.14	12.05	12.05	10.43
TV Connected Internet Device, Total App/Web on a Tablet	15.98	15.06	13.18	10.72

In this model, we assumed that "Total app/web on a smartphone" equals "video conferencing" due to a lack of representation of the "video conferencing" activity.

We used the same method for calculating the age group with a slightly modified equation:

$$\vec{X} = \langle \sum_{1}^{i=4} (\vec{I}_{i,j} \times A_i) \rangle$$

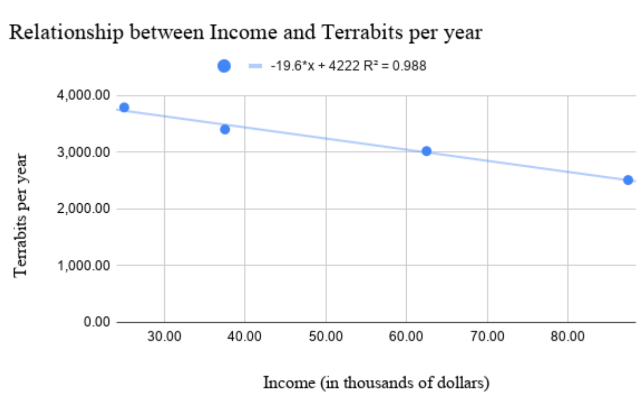
Using our data values, our new vector was:

$$\vec{X} = \langle 120.54, 108.21, 96.06, 79.84 \rangle$$

Then we found $\vec{X}_f$ by using dimensional analysis, with the same conversion as for the age category:

$$\vec{X}_f = \langle 3790.81, 3403.11, 3020.97, 2511.02 \rangle$$

Finally, we constructed the linear regression for income in thousands of dollars versus terabits per year.



Our R^2 value of .988 indicated a far better linear fit than for the previous regression. Thus, in cases where incomes are clear, the model used the income equation, but in cases where the individual was unemployed or income was not clear, the model used the age equation.

We then proceeded to find the total internet needs for 3 scenarios:

1. A couple in their early 30s (a teacher and someone who is looking for work) with a 3 year old child
2. A retired woman in her 70s who cares for 2 school-aged grandchildren twice a week receives a federal government pension of $22,172 [11]
3. 3 former M3 Challenge participants sharing an off-campus apartment, completing their undergraduate degrees, and working part-time make $36,824 [12]

For the first scenario, we assumed that the teacher earns $61,730 per year (the average salary of a teacher) [13]. For the person who is looking for work, we defaulted to the average bandwidth values found for the age categories. We used the same assumption for the 3 year old child. The minimum bandwidth necessary for 90% of the time was:

$$H = 0.9[-19.6(61.730) + 4222 + 4630 + 4630]$$
$$H = 2710.88 Tb/year$$

The same process was done for 99% of the time where we calculated that the couple with a child would need 2981.97 Tb/year.

For the second scenario, we assumed that the retired woman in her 70s earned $22,172 because of federal pension [11]. We also assumed that 2 children were in the 12-17 age category. We got the following for minimum H 90% of the time:

$$H = 0.9[-19.6(22.172) + 4222 + 4420 + 4420]$$
$$H = 11364.7 Tb/year$$

We did the same calculations 99% of the time, and calculated H = 12501.2 Tb/year.

For the third scenario, we assumed that a working part time undergraduate makes $36,824 [12]. Since there are 3 working students, we multiplied the individual internet needs by 3.

$$H = 3[-19.6(36.824) + 4222] \times .9$$
$$H = 9450.67 Tb/year$$

The minimum H 99% of the time was 10395.7 Tb/year.

Strengths and Weaknesses

In our model, we used sources that did not consider the impacts of the pandemic such as a shift in the work from home and online education. In addition, our model has limited flexibility in that we only take into account the households' income, excluding cases where individuals were unemployed or the income was unknown where we used age as a determination of minimum bandwidth. Additionally, our data may have been more accurate if we had more data from other years that was available for usage.

However, one strength of our model was the high correlated line of best fit in our income vs internet need with an R^2 value of 0.988. Additionally, we proved that age was not a good indicator of internet need; the income model proved it was a far better indicator of internet need with a high R^2 value. Finally, our models took into account multiple activities which strengthened it by utilizing multiple conditions.

Part III: Mobilizing Mobile—Mobile

Restatement of the Problem

We are tasked with constructing a mathematical model that determines an optimal plan for organizing the placement of cellular nodes within a region.

Assumptions

1. All three cell towers are the same price.
    - **Justification**: We used the average cost needed to deploy a 4G LTE tower in the US, which was $138,000 [14].
2. The population of each subregion is evenly distributed.
3. The data provided for regions A, B, and C are from the year 2020.

Symbols Used

D - the difference between the total cost of implementing a tower vs. the value given by the tower

n - the number of towers needed in a subregion based on A the square area of the subregion and r_i the range of the ith type of tower.

c - the cost of implementing towers in a subregion based on n, p the price of a cell tower, $b_d(x)$ the value of x the income of a subregion mapped to the bandwidth regression from part II

v - the value given by a cell tower within a single month based on n, h the number of households in the subregion, $\bar{p}$ the 2020 average price per Mbps of download speed, and b_i the bandwidth given by the ith tower type

Solution and Results

We chose to take the difference D between the total cost of implementing the three different types of towers c with the value given by all cell tower within a month v. A lower D value means that the cell tower performs better as the value v given is greater than the cost c.

We defined the number of towers needed n as the square mileage A of the region divided by the range r of the ith tower as a circle rounded up:

$$n = \frac{A}{\pi r_i^2}$$

We calculated the cost c as the sum of the number of towers needed to cover the region multiplied by the price p of each tower type and the minimum bandwidth in terms of Mbps of download speed needed $b_d(x)$ per household given the income x:

$$c = n \times p \times b_d(x)$$

We calculated the total value v that the the cell tower(s) of the ith type within the region as the number of towers n multiplied by the 2020 price per Mbps $\bar{p}$ multipled by the bandwidth by download speed in Mbps b_i given by the ith type of cell tower for the h number of households it serves that have smartphone access within a subregion:

$$v = n \times \bar{p} \times b_i \times h$$

This results in the final equation:

$$D = c - v = npb_d(x) - n\bar{p}b_i h$$
$$D = n(pb_d(x) - \bar{p}b_i h)$$

The following are the values of all constants in the equation:

$p = \$138000$ - taken from [14]

$\bar{p} = \$0.174$ - computed in Part I

$r_1 = 15\ mi, b_1 = 140\ Mbps$ - average range and download speed per user of a single node of a low band tower respectively

$r_2 = 2.5\ mi, b_2 = 500\ Mbps$ - average range and download speed per user of a single node of a medium band tower respectively

$r_3 = .75\ mi, b_3 = 1500\ Mbps$ - average range and download speed per user of a single node of a high band tower respectively

When applying this formula we get the following results:

Difference Computed by Tower

Region	Tower 1 (Low)	Tower 2 (Mid)	Tower 3 (High)	Predicted Bandwidth Per Household (Mbps)
A-1	26026	936914	10410039	110.17
A-2	16917	609006	6766652	108.31
A-3	12848	462531	5139165	98.22
A-4	34222	1231973	13688423	106.24
A-5	7635	274845	3053800	108.63
A-6	41168	1482049	16466995	98.54
B-1	44570	1604503	17827468	65.6
B-2	37385	1345840	14953345	44.02
B-3	18025	648890	7209432	19.9
B-4	26151	941431	10460120	57.74
B-5	88879	3199630	35550697	60.38
B-6	26396	950252	10557987	35.86

Region	Tower 1 (Low)	Tower 2 (Mid)	Tower 3 (High)	Predicted Bandwidth Per Household (Mbps)
B-7	66932	2409529	26771807	45.47
C-1	-411	-14798	-164455	-5.54
C-2	1851	66653	740573	67.74
C-3	176	6325	70268	9
C-4	1920	69123	768015	40.98
C-5	831	29931	332559	32.76
C-6	-63	-2267	-25223	-1.04
C-7	2206	79405	882249	33.23

Strengths and Weaknesses

The model does not account for the individual geometry of the subregion as the range of each cell tower is circular and the subregions are irregularly shaped or rectangular. The model also does not take population density into account. In addition, the model does poorly as income becomes too high, likely because it relies on the regression from part II to determine the expected bandwidth usage per subregion. Another limitation of the model was that we only focused on how many cell towers would be needed within a particular subregion instead of region itself, meaning that we did not produce an estimate of the best locations to place the cell towers nor does it attempt to maximize the Mbps given out to each household. Furthermore, the model only considers the value the cell tower provides within the next month, and would likely have produced better results if a longer time period was used.

However, the model determines which cell tower would be the best to use within a particular subregion, and according to the model, tower 1 (the low band tower) and works to suggest the best tower to place in a subregion so that it provides the minimum coverage needed for the particular subregion while minimizing cost versus value.

Bibliography

[1] https://arxiv.org/abs/2004.06098

[2] https://dl.acm.org/doi/10.1145/3419394.3423658

[3] https://www.jmir.org/2020/8/e20073/

[4] https://www.ncta.com/whats-new/report-where-does-the-majority-of-internet-traffic-come#:~:text=The 2019 Global Internet Phenomena, more video is being consumed.

[5] https://www.verizon.com/info/definitions/bandwidth/

[6] https://www.pewtrusts.org/en/research-and-analysis/articles/2020/11/30/how-much-broadband-speed-do-americans-need

[7] https://www.broadbandsearch.net/blog/internet-costs-compared-worldwide#post-navigation-3

[8] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1550668/

[9] https://telsoc.org/journal/ajtde-v4-n4/a67

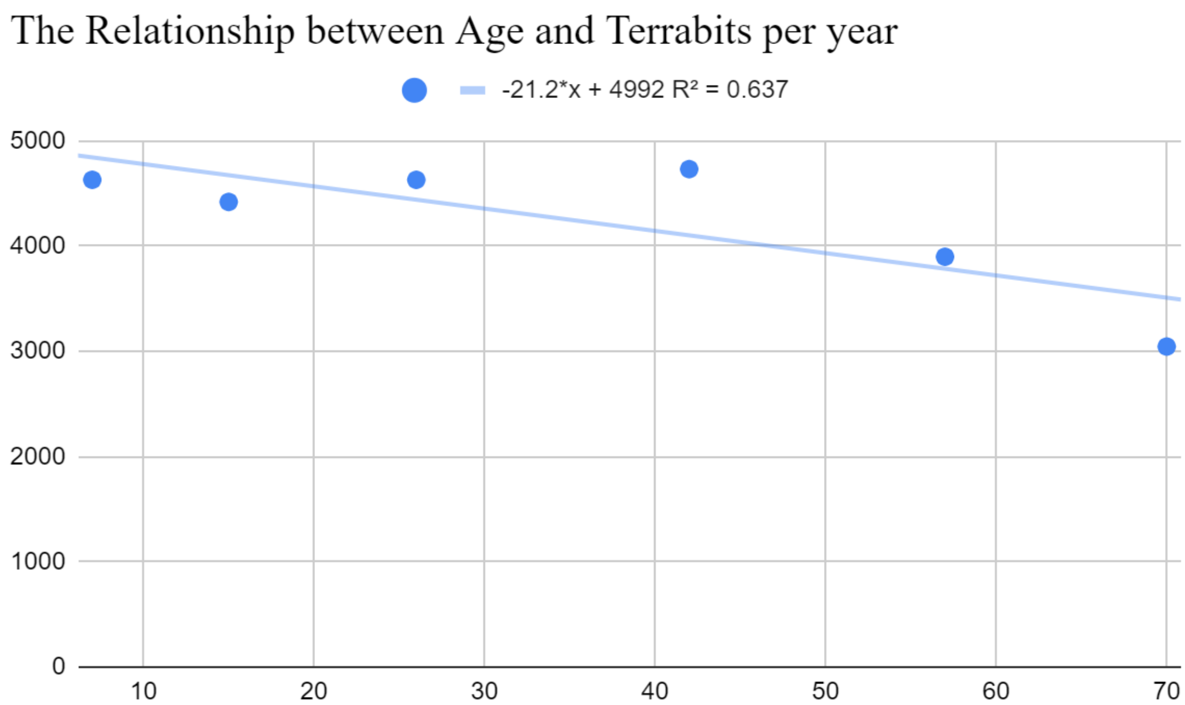
[10] https://www.npr.org/sections/health-shots/2016/01/14/462816458/average-age-of-first-time-moms-keeps-climbing-in-the-u-s

[11] https://www.personalcapital.com/blog/retirement-planning/average-retirement-income/#:~:text=Pension — Less than one-third,%2424%2C592 for a railroad pension

[12] https://www.ziprecruiter.com/Salaries/Part-Time-College-Student-Salary#:~:text=As of Feb 21%2C 2021,%2Fweek or %243%2C069%2Fmonth

[13] https://www.businessinsider.com/teacher-salary-in-every-state-2018-4#:~:text=Nationwide%2C the average public school,National Center for Education Statistics

[14] ce.toronto.edu/ychen/Materials/papers/infocom15.pdf


The Relationship between Age and Terrabits per year
-21.2*x + 4992 R² = 0.637
5000
4000
3000
2000
1000
0
10
20
30
40
50
60
70