Defeating the Digital Divide

Executive Summary

As technology becomes increasingly prevalent in our daily lives, especially during this unprecedented period of isolation due to the COVID-19 pandemic, we are forced to reevaluate the current state of connection. What was once face-to-face connection has been largely influenced by digital connections. In the developing world, it is more than important that the Internet, which connects people from all over the globe, is accessible to all.

In many third world countries, this has not been achieved, and even in countries such as the United States, digital connection is not always readily accessible, leading many to experience many disadvantages whether it be in planning for their future, accessing necessary educational structures, and communicating with essential services such as mental health lines or hospitals. It is crucial to make sure that the Internet is easily accessible to everyone. Although this barrier may not apply to everyone, it is undoubtedly an important issue to consider. In the following paper, we propose a variety of models to, in a way, combat this so-called "digital divide."

By modeling the internet service provider industry as ideally competitive [4], we narrowed down the factors to monthly price, average peak download speed (bandwidth) and time. A 3-dimensional planar regression on data from the US revealed a pattern that related years past 2020 as a function of speed and price. By manipulating the equation, price per unit was found to be constant over time. The methodology was immediately applied to data from the UK to make a corresponding model.

Right now, mobile broadband is needed for immersion in the digital world. Indeed, over the past several years, monthly internet demand in the United States has grown by a steady 24.1% annually [9]. A times series plot then allows us to predict growth in future years, assuming this constant growth holds. With the many usages of bandwidth for entertainment, work, education, etc., it becomes imperative to estimate with realistic models the minimum necessary bandwidth per household to ensure a productive environment.

Currently, mobile broadband technology pertains to Internet-enabled mobile devices such as smartphones and tablets. Internet connection is provided through mobile phone towers, which is then used by these devices [13]. In this new age of connection, mobile broadband technology provides unprecedented wireless Internet access. While we typically must be near a WiFi hotspot to receive Internet, mobile broadband technology allows us to access the Internet wherever we have cellular service, eliminating the need for WiFi networks and offering connection to the Internet in various locations [12]. To do so, we need to optimize the location of nodes in order to help make the Internet accessible for all. We found that through our modeling, a logical approach as to grouping subsections of different areas can create a map of the optimal node positions in order to create cheap and available Internet access for everyone.

The many possibilities for improving digital connections provides an avenue for our future development. Harnessing those possibilities will be the future not only of technology, but of connections.

1. The Cost of Connectivity

1.1 Restatement of the Problem

Develop a model to predict the cost per unit of bandwidth in dollars/pounds per Mbps over the next 10 years for consumers in the United States and United Kingdom.

1.2 Assumptions

1. **The overall number of households with internet connections will remain unchanged.**
   a. Internet connection is essentially a household necessity [1], or in economic terms, an almost perfectly inelastic product. Therefore, one can assume that changes in price level do not impact the overall quantity demanded, and modeling should not have to consider consumer entry/exit from the internet connection market. However, changes in type of internet connection (speed) due to price can still exist.
   b. The logic is that as a necessity, every household that requires an internet connection will already have it by now, and the rare households that do not have little incentive to buy it. However, households that do may choose to upgrade their wifi based on their needs.
2. **Price level of unit of bandwidth is determined by aggregate demand.**
   a. The internet provider market functions as a near-perfectly competitive oligopoly, as high infrastructure costs prevent entry/exit, product (internet) is standardized, and there are only a few firms in the industry [4], yet they are competitive to beat others prices [6].
   b. Competitiveness means that the market sets the prices, not the vendor, and in the long run (where the supply curve is vertical) price is determined by shifts in the demand curve. Thus, the price level is dependent on aggregate demand, which, due to assumption 1, will be a function of price and speed (or utility).
   c. Note aggregate demand is the demand of the market overall, not a singular firm.
3. **Aggregate demand is a function of expectations.**
   a. Economic theory states that shifts in aggregate demands come from change in tastes, number of buyers, price of substitutes/related goods, income, and consumer expectations [4]. Internet connection's status as a necessity holds the first three factors as constant. Predicted change in income is not within our scope and will not be considered a factor. However, as the internet grows more and more involved with daily life, consumer expectations are shifting demand up, where it becomes as common as electricity [8].
4. **Price and bandwidth speed are linearly related.**
   a. Since infrastructure is already laid out and Assumption 2 states it is a competitive market, price will be proportional to marginal cost of providing an additional Mbpc [4].
5. **Relationships between time, price, and speed are similar between the US and the UK.**
   a. Due to the free market economy being similar in most developed nations [4], previous assumptions will hold true and relationships derived from them will be as well.

1.3 Model Development

1.3.1 Variables

Symbol	Name	Units	Value
Y	Time in years since 2000	Years	Current year - 2000
P_1	Monthly price of broadband in US	Dollars	
P_2	Monthly price of broadband in UK	Pounds	
P_{US}	Unit price (US)	Dollars / Mbps	P_1 / S
P_{UK}	Unit price (UK)	Pounds / Mbps	P_2 / S
S	Average Peak Download Speed	Mbps	
S_1	Modified S_1 for US	Mbps	See below

Table 1.1

1.3.2 Methods

Unit price of bandwidth, at the very basic level, is set by internet service provider companies. The overall question becomes how do trends over time in the broadband industry impact the companies' decision on what price to put on their product. According to microeconomic theory, interactions between what we assume are rational consumers and companies (they are maximizing utility with their resources) should be predictable.

Assumption 1 states that due to the necessity-like status of some sort of internet connection, we can assume that every household that has internet now will keep some level of internet connection in the future. However, price levels and speeds can impact that level, as higher prices may force households to purchase worse plans but higher broadband speeds may incentivize them to purchase better plans. Overall, this prevents having to consider effects on price on amount of people getting internet at all, which could be messy, however we still must look at price and speed trends over time.

Assumption 2 states that unit prices are a function of the aggregate demand for them only, which in turn Assumption 3 relates to expectations. As stated in Assumption 3, expectations will increase over time, making unit prices a function of time. This relates back to the original problem, where we need to build a model for unit price of bandwidth over time.

Through these assumptions, we have isolated three variables. Since unit price is a function of internet speed and price (where unit price equals price over speed, see Table 1), we consider internet speed and price. We will also be analyzing time. Linear regressions were done with data from Defeating the Digital Divide Data, MathWorks Math Modeling Challenge 2021 to determine the type of relationship each variable has with each other. Additionally, linear regression t-tests were done to determine if each pair of variables had a statistically significant relationship ($p < 0.05$) , as shown below:

$$P_1(S) \sim aS + b$$

Year	a	b	r	p

2012	0.37090	54.0679	+0.9393	0.00541
2020	0.04185	46.75091	+0.7262	0.00217

Table 1.2

While the average speed is not representative of the rest of the US, Assumption 4 lets us believe that price should scale down with speed when this overinflated average speed is brought down to the true average. Overall, the strong r values and low p values prove a relationship and linearity.

Modeling internet speed versus years is more complicated due to a data hole. The methodology used to collect the data was changed in 2017 (Y=17) with both methodologies overlapping to create one dual 2017 entry. The linearity of both were modeled separately:

$$S(Y) \sim aY + b$$

2010 - 2017 (inclusive)				
Country	**a**	**b**	**r**	**p**
US	9.5750	-85.4250	+0.9735	4.569e-05
UK	8.9905	-81.3214	+0.9920	1.294e-06

Table 1.3

The data provider switches at 2017, and due to the difference in methodology the data cannot be accurately modeled together. To ensure linearity, we also modeled the data from the next provider from 2017 through 2021 (there is an overlap for 2017). These results are shown below:

2017 - 2021 (inclusive) -				
Country	**a**	**b**	**r**	**p**
US	25.740	-374.271	+0.9862	0.00195
UK	7.393	-79.093	+0.9334	0.02043

Table 1.4

Additionally, the similar trends in the US and UK help provide empirical evidence for Assumption 5. Working off of that, we modeled the relationship between time and monthly price using data from the UK, as shown below:

$$P_2(Y) \sim aY + b$$

a	**b**	**r**	**p**
-1.3202	33.0652	-0.7922	6.641e-06

Table 1.5

From the relationship analysis, all r values are strong, indicating good linear correlation. Linear models are able to model this data well, which is important for further modeling. Additionally, each variable is shown to have a significant correlation with each other due to p values < 0.05.

1.4 Model Execution

So far, we have modeled monthly price, average peak download speed, and time in years, seeing how each linearly relate to each other. Since they were done using different datasets and different methodologies, we cannot construct them into a master data set using the linearity equations, but instead use it as a guiding principle to create a new model.

Set up a 3-dimensional (S_1, P_1, Y) space. The $S_1 - P_1$, $S_1 - Y$, and $P_1 - Y$ planes now represent a space where the linear relationships can be expressed as lines. laying on their respective planes. We then modeled year, price, and internet speed together as a 3-dimensional object with three linear traces, essentially, a plane.

We created a planar regression using data from Defeating the Digital Divide Data, MathWorks Math Modeling Challenge 2021. Since the distribution of average peak download speed and average monthly prices are higher than representative of the entire US, we needed to scale it down to the modified speed S_1 and price. We did this by comparing it to previously used data that records the US average peak download speed at 134.8 Mbps (the unrepresentative sample average is 482 Mbps). Thus each speed S was scaled down to modified speed S_1 by:

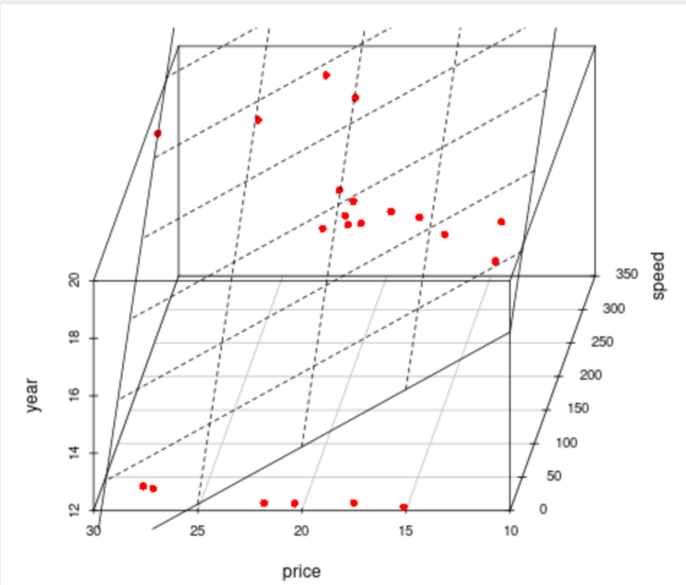
$$S_1 = \frac{134.8}{482} S$$

Due to Assumption 4, the price is also scaled down by the same factor. The resulting data is below:

Y	S	P	S_1	P_1
20.00	313.35	100.00	87.63	11.43
20.00	787.53	200.00	220.25	29.47
20.00	974.53	250.00	272.55	20.62
20.00	337.76	100.00	94.46	15.45
20.00	858.18	225.00	240.01	24.89
20.00	279.01	125.00	78.03	19.90
20.00	298.88	100.00	83.59	18.76
20.00	1096.38	300.00	306.62	22.41
20.00	103.83	42.00	29.04	11.04
20.00	248.00	72.50	69.36	13.94
20.00	305.48	200.00	85.43	18.16
20.00	370.27	100.00	103.55	16.93
20.00	423.92	200.00	118.56	18.91
20.00	482.94	300.00	135.06	19.78
20.00	348.13	150.00	97.36	19.05
12.00	129.00	25.00	36.08	28.03
12.00	119.00	30.00	33.28	27.53
12.00	41.00	33.00	11.47	21.95

12.00	19.00	16.50	5.31	15.16
12.00	37.00	20.00	10.35	20.47
12.00	41.00	30.00	11.47	17.64

Although there are only two points for Y, we can assume that the model is linear with respect to Y due to the work done in proving linearity previously. Planar regression of the data was done in R, and resulted in the following model:




```
Call:
lm(formula = year ~ speed + price, data = dataset)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.3292 -1.8192  0.8575  1.4479  3.1750 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 22.210830   2.139128  10.383 4.99e-09 ***
speed        0.032856   0.006312   5.205 5.97e-05 ***
price       -0.399478   0.111845  -3.572  0.00218 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.363 on 18 degrees of freedom
Multiple R-squared:  0.6334,	Adjusted R-squared:  0.5927 
F-statistic: 15.55 on 2 and 18 DF,  p-value: 0.0001195
```

Figure 1.1

From this data in Figure 1.1, a model can be created that creates a speed-price relation for any year:

$$Y(S, P) = 0.032856(S) - 0.399478(P) + 22.210830$$

For example, $Y(S, P) = 31$ would create a speed-price line exactly 10 years from now in 2031, when graphed on the S-P plane. However, in determining unit price, further manipulation is needed. Remember that unit price is essentially price per unit of bandwidth, or price per Mbps (the slope of the speed-price line in (S, P) space). We find slope through implicit differentiation:

$$\frac{dP}{dS} = \frac{0.032856}{0.399478} = 0.08225$$

Note that unit price does not change with time. This model concludes that unit price is a constant \$0.08 USD per Mbps in the US. With Assumption 5, we can state that the same conclusion will be reached in the UK. While we do not have a dataset will all three values for UK, we can simply calculate the range of values: where price ranges from £17 to £26 and speed ranging from 0 to 30 Mbps, we can calculate an average of £0.88 Mbps, which we can then assume is constant.

1.5 Discussion

This model is supported as it quantifies certain relationships. For example, one can see how the variables relate to each other by finding the gradient of the plane:

$$\nabla Y\left(S, P\right) = \langle 0.032856, -0.399478\rangle$$

As Y increases, all else being equal, S increases, as technology improves and both supports and requires faster internet speeds. This is corroborated by the strongly positive r values from Tables 3 and 4. As Y increases, all else being equal, P decreases, as the constant internet speed suffices less and less for the growing bandwidth demands of technology and is therefore worth less. This is also matched by the strongly negative r value from Table 5.

This model predicts theoretical trends in economics quite well, and as shown in the planar regression, it carries over to empirical evidence. The strength in this model lays in its versatility, mainly its ability to create speed-price relationships every year. Looking at the overall R-squared value, 0.6334, the model is fairly strongly correlated as well.

The primary weaknesses of this model would likely come from assumptions. The main basis of the model was to be able to not account for market entry/exit as it was built in a fully built infrastructure ideal economic situation. However, numerous factors allow a less than ideal situation, such as collusion between internet giants, currently expanding fiber infrastructure, and current economics problems forcing homes to give up internet connection. Due to its basis on economic theory, it cannot handle many deviating factors.

1.5.1 Sensitivity Analysis

Consider the model of the years since 2000 function for the US:

$$Y\left(S, P\right) = 0.032856\left(S\right) - 0.399478\left(P\right) + 22.210830$$

Due to the low coefficients of the model with respect to S and P relative to the constant, it will likely not have too much change. This can be proven:

$$\frac{\partial Y}{\partial S} = 0.032856 = 3.29$$

$$\frac{\partial Y}{\partial P} = 0.399478 = 39.95$$

For every one Mbps change in speed, years change by 0.033. For everyone $1 change in price, years change by 0.4. This is a very low change, and therefore the models predictions are more likely accurate.

2. Bit by Bit

2.1 Restatement of the Problem

Create a flexible mathematical model to predict a given household's need for the internet over the course of a year. Apply your model to 3 example households and determine the minimum amount of required bandwidth that would cover their total internet needs 90% of the time and 99% of the time?

2.2 Primary Assumptions

1. The distribution of time online for each age group is the same in Q1 of 2019 and Q1 of 2020.
    a. Using a chi-squared test for homogeneity for the distribution of weekly time online in Q1 of 2019 and 2020 with age group [16], we concluded that there is not statistically significant evidence to suggest a different frequency in these 2 years (df = 5; p = 0.9995; Fig 1). This suggests that the proportion of time that Americans of each age group spent online remained the same in Q1 of 2019 as in 2020.

Contingency Table (Observed Frequencies / Expected Frequencies):

	Q1 Use		
Age Group:	2019 Q1	2020 Q1	
2-11	22 (21.25)	22 (23.30)	45
12-17	16 (15.54)	16 (17.04)	33
18-34	56 (56.39)	62 (61.86)	118
35-49	69 (68.80)	76 (75.47)	144
50-64	75 (75.95)	84 (83.31)	159
65+	77 (76.94)	85 (84.38)	161
	315	345	660

Inferential Statistics:

Chi Square:	0.1575
df:	5
p (χ^2):	0.9995

    b.

Figure 1

2. Monthly internet usage in the United States holds constant between January and September of the years in our model
    a. With a chi-squared goodness-of-fit test for the internet usage per month from January to September 2019 [5], we found that there is no statistically significant evidence to suggest that the monthly internet usage in the United States changes from January to September 2019 (df = 8, p = 1.0000; Fig. 2), leading us to reasonably assume that this pattern holds for future years as well.

b.

Goodness of Fit Test

Descriptive Statistics (Variable: Internet Usage):

Category	Observed f (p)	Expected f (p)	$(O-E)^2/E$
Jan	270 (0.1098)	273.3356 (0.1111)	0.0407
Feb	273 (0.1108)	273.3356 (0.1111)	0.0026
Mar	275 (0.1118)	273.3356 (0.1111)	0.0101
Apr	274 (0.1115)	273.3356 (0.1111)	0.0025
May	273 (0.1111)	273.3356 (0.1111)	0.0000
Jun	273 (0.1108)	273.3356 (0.1111)	0.0026
Jul	273 (0.1111)	273.3356 (0.1111)	0.0000
Aug	274 (0.1115)	273.3356 (0.1111)	0.0025
Sept	275 (0.1118)	273.3356 (0.1111)	0.0101
Total:	2460	2460.0200	0.0712

Inferential Statistics:

Chi Square:	0.0712
df:	8
p (x^2):	1.0000

Figure 2

3. The internet traffic in the United States is assumed to grow at a rate of 24.1% each year.
   a. From 2018 to 2020, the internet traffic in the United States (in million exabytes per month) has grown by 24.1% each year [9]. Thus, this pattern is expected to hold for future years as well, up to 2023 (disregarding effects of COVID-19, whose data has not yet been published). As such, we made the reasonable assumption that there is a 24.1% increase every year in internet traffic during the period 2018 to 2023.
   b.

X ×

2x6 double

	1	2	3	4	5	6
1	2018	2019	2020	2021	2022	2023
2	33.4500	41.5200	51.5500	64	79.4600	98.6400

   c.

X × percentIncrease ×

2x5 double

	1	2	3	4	5
1	2019	2020	2021	2022	2023
2	0.2413	0.2416	0.2415	0.2416	0.2414

4. Increased internet usage leads to increased probability of concurrent device usage.
   a. This assumption makes intuitive sense. As demand for internet increases, it can be reasonably assumed that this phenomenon is partly caused by the occurrence of more individuals in a household accessing the internet with separate devices.

2.3 Developing the Model

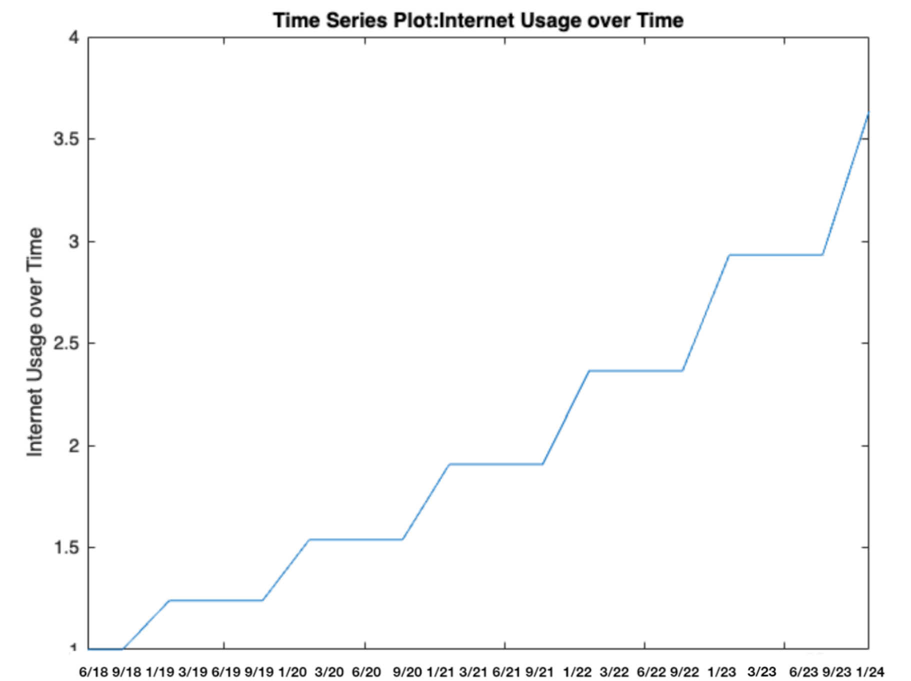
1. Time Series of Internet Demand
    a. To create a graphic display that shows demand for internet usage over time, we created a time series plot in MATLAB that predicts demand relative to June 2018 (268 GB per month), which was the most recent data we could find [10]. By Assumptions 2 and 3, the demand for internet increases by 24.1% annually, with demand remaining constant from January to September of each year. We further assume that future demand displays linear growth from September to January, as this was the trend during 2018 and 2019 [5].
    b. By Assumption 4, the predicted trend of internet usage will correspond to the likelihood of concurrent device usage. As such, this trend prediction shown in the time series plot sheds light on the number of concurrent devices accessing the internet at given time periods throughout the day, thus allowing for a prediction model of minimum necessary bandwidth for households.
    c. MATLAB code included in appendix.
    d. 

Figure 4

2.4 Model Execution

The model construction is based on conditionals and the inputs of the user for determining the minimum bandwidth required. The model constructed must be flexible, being able to take into a lot of different variables (e.g. different age groups, remote or in-person working, etc.) and situations. It is important to note that the bandwidth is additive meaning that it does not depend on the amount of time but rather depends on the amount of devices and/or activities being done at the same time.

2.5 Assumptions

1. Internet usage times are assumed to be the following periods for the following age groups and type of digital access.
    a.

			Usage times of activities					
	Age Category							
Activity	**2-11**	**12-17**	**18-64 (H)**	**18-64 (W)**	**65+**	**Daily usage frequencies**		
Watching Traditional Television						ONCE		
TV Connected Game Console	3pm - 9pm	3pm - 9pm	5pm - 12am *	5pm - 12am *	8am - 10pm	ONCE		
TV Connected Internet Device**	3pm - 9pm	3pm - 9pm	5pm - 12am *	5pm - 12am *	8am - 10pm	ONCE	(H) work at home / not working	
Internet on a Computer (not including video)			8am - 10pm	5pm - 12am *	8am - 10pm	MULTIPLE	(W) work away from home	
Video on a Computer			5pm - 12am *	5pm - 12am *	8am - 10pm	ONCE	* denotes the following day	
Total App/Web on a Smartphone			8am - 12am *	5pm - 12am *	8am - 10pm	MULTIPLE		
Video Focused App/Web on a Smartphone				5pm - 12am *	8am - 10pm		for (H)	Anything below 0.16 hours (16 * 0.01) is negligible
Steaming Audio on a Smartphone				5pm - 12am *	8am - 10pm		for (W)	Anything below 0.07 hours (7 * 0.01) is negligible
Total App/Web on a Tablet			8am - 12am *	5pm - 12am *	8am - 10pm	MULTIPLE		
Video Focused App/Web on a Smartphone				5pm - 12am *	8am - 10pm			
Steaming Audio on a Smartphone				5pm - 12am *	8am - 10pm			

Table 1

    b. The workday is assumed to be from 8am to 5pm for all adults working outside the home. With an assumed bedtime of 12am, this provided the usage times shown for this group.
    c. If an individual is looking for work or is working from home, then they use the internet from 8am to 12am the next day (i.e. total time awake), with entertainment (videos, TV) from 5pm to 12am.
    d. For school age children (ages 2-11 and 12-17), it is assumed that they will be at school from 8am to 3pm, leaving their usage times to be from 3pm to 9pm.
    e. For individuals aged 65+, it is assumed that they are retired and will remain home for the majority of the day, giving their usage times to be from 8am to 10pm (i.e. times they are assumed to be awake).
2. 35% of the advertised bandwidth is lost before usage by devices
    a. The average WiFi speed in homes will generally be 20% to 50% less than the advertised maximum download speed, due to wireless interference and fade [11]. 35% (the midpoint) is used as a benchmark for the purposes of this model.

Activity	Actual Needed Bandwidth
General web surfing, email, social media	1.35
Online gaming	2.7
Video conferencing	3.375
Standard definition video streaming	4.725
Minimum *required* speed for Netflix video streaming.	0.675
Minimum speed for SD streaming with YouTube.	0.945
Netflix *recommendation* for SD video streaming.	4.05
High definition video streaming	8.775
Minimum speed for HD streaming on YouTube.	3.375
Netflix *recommendation* for HD video streaming.	6.75
Netflix *recommendation* for Ultra HD (4K) video streaming.	33.75
Frequent large file downloads	67.5

Table 2

3. Traditional television is excluded from the model.
    a. Most traditional TVs (i.e. cable TV) require no bandwidth access.
4. Streaming on TV is assumed to be Netflix HD streaming, and streaming on mobile device is assumed to be YouTube streaming

2.6 90% And 99% Bandwidth Requirement

An individual that works from home will be having access to the internet throughout the day thus 16 hours will be available. A working individual that does not work from home will be having household internet access for just 7 hours. With the 99% model, 1% of the time will be ignored. This 1% will exclude time periods where there is a maximum bandwidth usage. The light blue blocks are the activities that can be ignored by this 99% model due to the fact that they will either be out bandwidthd no matter what or their time period is too small so that the 1% time excluded will cut them out. If there is an individual at home then it will be 0.16 hour cut off and if there are no individuals at home then it will be 0.07 hour cut off.

The 90% necessity can be done similarly, where 1.6 hours and 0.7 hours are excluded for at home and only at work individuals. The model will be similarly constructed so that there is maximum bandwidth usage. This becomes more complicated as some activities will take more time than what is excluded. So the model must attempt to exclude as many high bandwidth activities in the given time period.

SITUATION 1: A couple in their early 30's (one is looking for work and the other is a teacher) with a 3-year-old child.

The model is applied to obtain a value of 13.5 Mbps bandwidth as the minimum requirement for 99% of the time during the day. 99% of the day will exclude low time usage activities from the model. The exclusion depends on the total home internet usage though, which depends on whether there is an adult working from home or not. In the case that an adult is working from home (or is unemployed and at home), then the amount of time that will be excluded from the model is 16 * 0.01 hours which is equivalent to 0.16 hours. If no one is at home, then the excluded time will be 7 * 0.01 hours which is equivalent to 0.07 hours.

The first step in determining the minimum bandwidth required is to determine a time interval where the most bandwidth is being used across the most possible amount of devices. This will likely happen when there is the most amount of streaming due to the fact that streaming takes up the most bandwidth. The 3 year old is the most limited in device use and thus the model will use the TV streaming (for 0.793 hours) for this individual which adds 6.75 Mbps bandwidth. Now the two adults in their early 30's must be taken into account for. What makes this tricky is that the adults will have different household internet usage due to the fact that one is employed and the other is. The unemployed individual will have an internet use time period from 8 AM to 12 AM the next day, while the employed teacher will have an internet use time period from 5 PM to 12 AM the next day. This means that the amount of household internet use for the unemployed individual will be more distributed throughout the day. But for the sake of finding a minimum bandwidth requirement, the model must maximize the amount of device/activity overlap. This means that both individuals will use either their tablet or their computer during the time that the 3 year old is watching TV. We assumed that streaming on youtube in HD is 3.375 Mbps. The final total for this situation will be equal to 6.75 + 3.375 + 3.375 = 13.5 Mbps.

Additionally, the probability that the maximum bandwidth is needed can be calculated. This can be done by using the number of hours daily and finding the intersection of all the individuals using their high bandwidth activities at the same time. Furthermore, an analysis of the bandwidth over the duration of the day can be done with construction of probability functions. Neither of these were unable to be done during this project due to the time constraint.

SITUATION 2: A retired woman in her 70's who cares for two school-aged grandchildren twice a week.

This can be done similarly to Situation 1. School-aged is undefined by the prompt, but we assume that it is for the age group of 12 - 17 years old. We can not combine it with 2 - 11 years old due to the fact that these two age groups have different TV time between streaming and as a game console. The fact that the children come only for two days is irrelevant as we have to find a maximum bandwidth which is when all three individuals are doing activities. It is important to note that since there are two grandchildren, only one of them will be doing an internet use activity due to the fact that our model only takes into account one TV. The minimum required bandwidth will likely be much lower than in situation 1 due to the one TV limitation. This situation was not modeled due to the lack of time.

SITUATION 3: Three former M3 Challenge participants sharing an off-campus apartment while they complete their undergraduate degrees full-time and work part-time.

This situation is harder than the last 2 situations due to the fact that the internet use as a college student and as a working individual must be taken into account. Their internet use can vary dramatically depending on many factors such as working from home or studying from home. Along with this, it is likely their work schedules are different resulting in maximum bandwidth usage at more obscure timings if possible.

2.7 Limitations

1. We assume that bandwidth usage during weekends is the same as weekdays.
2. In our model, only one person can use the TV at any given time.
3. This model does not take in account other types of internet usage activities other than the ones given in the provided dataset (D4).
4. This model assumes homogeneity within all members of each age group. In reality, this is not the case, as Americans come from varying economic and societal backgrounds, and thus will display varying trends in internet usage.
5. A significant limitation of this model is that it does not take into account the effects of the COVID-19 pandemic on bandwidth demand. Indeed, the pandemic has led to increased amounts of Americans working or learning from home. Had we had more time, we would have researched the effects of COVID-19 on internet demand further, and provided an alternate model that takes into account the unprecedented event, instead of assuming constant 24.1% annual growth in internet demand (as in Primary Assumption #3).

2.8 Time Series Application

Two models were generated for this problem. One is the time series and one is for determining bandwidth requirements. These models go hand in hand. The problem at hand was to determine bandwidth over the course of a year and this is where the time series comes into play. Everything described for the bandwidth model was for within the duration of a day. The time series is used through the substitution with the date. Depending on the year, a base factor (y-axis) can be multiplied to the bandwidth.

Indeed, using the time series plot, not only can we predict the minimum bandwidth requirement for the current time period, but we can also predict for future years as well. Since internet demand is assumed to be linearly related with minimum bandwidth requirement (Primary assumption #4), increased internet demand results in high probabilities of increased concurrent device usage. And since bandwidth is additive (depending on the number of devices accessing the internet at any given time), by examining the time series plot generated in MATLAB, we can predict bandwidth usage for up until the end of 2023, where the time series plot begins to produce less confident results.

Had we had more time, we would have expanded on this application further. As it is, the current time series plot provides a simple track to predict bandwidth requirements for upcoming years.

3. Mobilizing Mobile

3.1 Restatement of the Problem

Where to place the radio towers is an important problem to solve in order to give widespread mobile broadband access as efficiently as possible to everyone in the world [12]. The bandwidth needs of different regions must be satisfied by these cellular nodes, and factors such as the cost and demographics must be considered as well. We aim to develop a logically based model to determine where to best place these cellular nodes.

3.2 Assumptions

1. People spend 2 hours and 51 minutes per day on their mobile device.
    a. Justification: The average person spends 2 hours and 51 minutes according to Milijic [14].
2. People use around 3 Mbps when using the Internet on their phones.
    a. Justification: According to NerdWallet, websites, email, and social media require 1 Mbps and video streaming requires around 6 Mbps. On average, most people familiar with the Internet will use 3 Mbps at a time [15].
3. Older people (> 60) tend to use less Mbps, around 2 Mbps when using the Internet on their phones.
    a. Older people watch fewer videos and stream less audio, thus lowering their mbps usage according to the Nielsen Corporation's report. Provided on the Information Sheet [16].

3.3 Variables

Symbol	Definition	Units	Value
n	Subregion number	No units	1 ~ number of subregions in a region
P_n	Population of each subregion	Number of people	See information sheet.
A_n	Area	Miles squared	See information sheet.
MA_n	Median age	Years	See information sheet.
T	Average person's time spent on a mobile device per day	Seconds of screen time per day	10260 s/day [14]

B_Y	Mobile usage by younger people (< 60) on average per second	Mbps	3 Mbps [16]
B_O	Mobile usage by older people (> 60) on average per second	Mbps	2 Mbps [16]
M_n	Total Mbps used in a subregion.	Mbps	…
R	Farthest distance from node to designated region.	Miles	…

Note: T was found by using Assumption 1 and converting 2 hrs and 51 minutes to seconds.

3.3.1 Developing the Model

We decided to employ a logical operator model in order to adapt the necessary distribution of node locations for different regions. Our primary goal is to get the entire population of each subregion access to the Internet through mobile broadband technology.

To do so, we consider three metrics:

1. Population and population descriptors
2. Area
3. Cost

For the first metric, we calculate M_n, or the total Mbps used in the subregions through the piecewise function. The goal of both *Eqn 3.1* and *Eqn 3.2* is to determine the total amount of Mbps used in a subregion by determining the mobile usage of each person (separated by the median age of a subregion) and then multiplying it by the population of the subregion:

M_n = (piecewise function)

If ($MA_n < 60$):

$P * T * B_Y$

If ($MA_n > 60$):

$P * T * B_O$

We then assign a bandwidth level for each subregion, making sure the typical download speed of each node is larger than the amount of bandwidth being used (M_n) at any given time:

$30 < M_n < 250$ => Low bandwidth

$100 < M_n < 900$ => Mid bandwidth

$1000 < M_n < 3000$ => High bandwidth

$3000 < M_n$ => Separate the subregion *[Eqn 3.3]*

Next, we want to see the area of the subregions and the area that nodes can cover. The area of the subregions are given in the information sheet [16]. We can also calculate the area that the nodes cover. Since we are given the radius for each node, we can use $A = \pi r^2$ to determine the range covered by a single antenna (see Figure 3.1).

Band	**Typical Download Speed**	**Range of Single Node/Antenna**	**Area Range covered by Single Node/Antenna**
Low	30-250 Mbps	10-20 miles (16-32 km)	314.159 to 1256.637 square miles
Mid	100-900 Mbps	2-3 miles (3.2 -4.8 km)	12.566 to 28.274 square miles
High	1000-3000 Mbps	.5-1 miles (.8-1.6 km)	0.785 to 3.142 square miles

Figure 3.1

However, we want to ensure that every region gets covered in order to ensure access at all times. Thus, we make sure that most of the area falls in the smaller range (as seen by the darker purple Figure 3.2). Although the lighter purple may get some connection, it is not entirely ensured and thus we try to avoid having regions covered only by the lighter section of a single node.

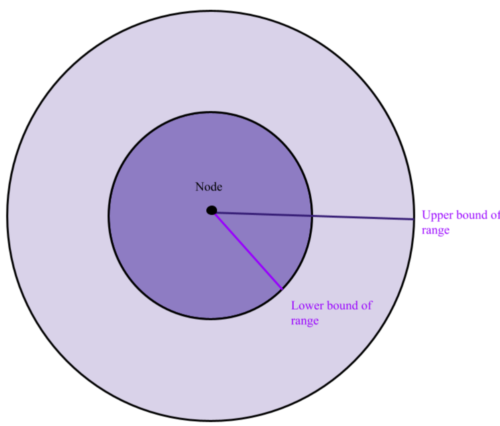


Figure 3.2

We check to see if our bandwidth level assigned covers all of the area. If it does not, then we split the subregion into two separate sections, each with a node that has a typical download speed larger than the amount of bandwidth being used (M_n) at any given time (see Equation 3.3) and a large enough area to cover all of the section.

Finally, the last factor we need to consider is cost. It is more beneficial to have fewer nodes in order to minimize the cost. Thus we combine adjacent nodes to a higher level (ie. low to mid) if the limiting factor is M_n and combine adjacent nodes to a lower level (ie. high to mid) if the limiting factor is area. To do so, we combine two adjacent sections and confirm that they work according to the first two parameters, population and area, without needing to be separated.

After sectioning out the subregions, we determined the centroid of each section to make sure we placed the node in a spot that reaches the bounds. We did this through Geogebra, where we traced out each section and found the location of the centroids (Figure 3.7 in Results and Discussion). Since the centroid is the geometric center of a figure and is the arithmetic mean of the section needed to be covered, it is the optimal place to put the node in order to ensure mobile broadband access to the entirety of the region.

3.4 Model Execution

Using *Region A* as an example, we calculated the total Mbps used in each subregion (M_n). *Equation 3.1* was used for all the subregions except for Subregion 4, which uses *Equation 3.2* because Subregion 4 had a median age (MA_n) greater than 60. Note in Figure 3.1 that Subregion 4's average mobile usage per second (B) is highlighted in green because it is a different value. After combining the subregions and choosing the appropriate Band level according to the process explained above in Developing the Model, we combined Subregions 1 and 2 with a Medium-level Band as the area of the two subregions summed to 2.01 square miles with a combined necessary download speed of 752 Mbps. We also combined Subregions 4 and 5, utilizing a Medium-level Band as the area summed to 2.01 square miles with a necessary download speed of 541.856 Mbps. We did not choose to combine Subregions 3 and 5 because their download speeds summed to greater than 900 Mbps, which exceeded the Medium-level Band range of download speed while having an area too large to match the High-level Band's lower range of area. Again, we wanted to ensure that the area would receive the necessary download speed, so because the area was greater than the High-level Band's lower area range, we did not choose a High-level Band. Thus, we chose to keep Subregions 3 and 6 separate, choosing a Medium-level Band for both. We then determined the centroid of the subsections of each combined and separate region, which can be seen in Figure 3.7.

n	Subregion 1	Subregion 2	Subregion 3	Subregion 4	Subregion 5	Subregion 6
population (P_n)	690	1422	1303	278	1243	1391
Time (T)	10260	10260	10260	10260	10260	10260
median age (MA_n)	28.2	30.2	40.7	64.3	37.8	36.9
Average mobile usage per second (B)	3	3	3	2	3	3
Total Mbps used in region (M_n)	245.8125	506.5875	464.19375	66.025	442.81875	495.54375
area (square miles)	1.21	0.8	0.67	1.65	0.36	2.14

Figure 3.3 *Region A*

n	Subregion 1	Subregion 2	Subregion 3	Subregion 4	Subregion 5	Subregion 6	Subregion 7
population (P_n)	3873	2114	1253	1129	2493	2398	1794
Time (T)	10260	10260	10260	10260	10260	10260	10260
median age (MA_n)	47.7	48.2	59	36.3	55.1	45.5	53.8
Average mobile usage per second (B)	3	3	3	3	3	3	3

Total Mbps used in region (M_n)	1379.75625	753.1125	446.38125	402.20625	888.13125	854.2875	639.1125
area (square miles)	3.48	4.35	4.64	2.32	7.54	3.77	7.54

Figure 3.4 *Region B*

n	Subregion 1	Subregion 2	Subregion 3	Subregion 4	Subregion 5	Subregion 6	Subregion 7
population (P_n)	1468	1624	1012	1295	1309	1008	1789
Time (T)	10260	10260	10260	10260	10260	10260	10260
median age (MA_n)	41.67	34.9	47.1	44.7	32.7	47.9	40.7
Average mobile usage per second (B)	3	3	3	3	3	3	3
Total Mbps used in region (M_n)	522.975	578.55	360.525	461.34375	466.33125	359.1	637.33125
area (square miles)	0.38	0.14	0.1	0.24	0.13	0.31	0.34

Figure 3.5 *Region C*

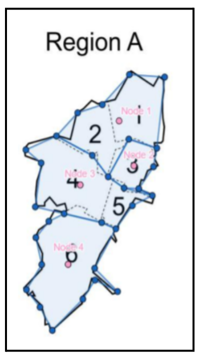


Figure 3.6 *Region A Example*

3.5 Results and Discussion

The results showed the locations of the nodes in order to minimize costs but still ensure complete coverage and enough bandwidth to each of the regions.

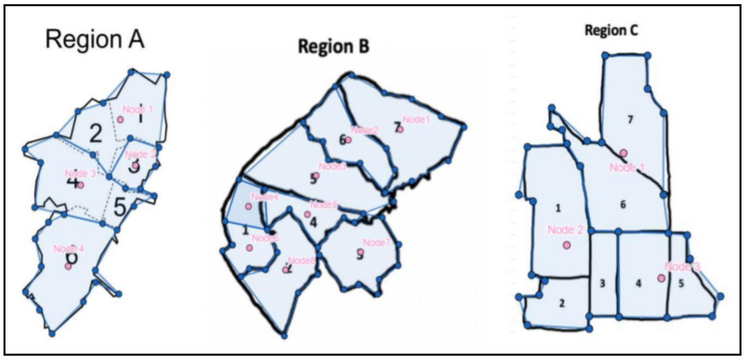


Figure 3.7 *Node Placements in Regions*

3.6 Strengths

This model's largest strength is that it adequately covers each region, focusing on placing the nodes within the distance needed to provide broadband Internet access. In this sense, it ensures that the entire population has access to broadband Internet service. The choice to combine subregions and lower the number of nodes also helps to reduce the costs of this model.

3.7 Limitations

We overcompensated for data usage in order to ensure that there is enough bandwidth. This is because we did not consider the percentage of households that already have access to mobile broadband Internet access [16].

In addition, to make sure that all of the area is covered, we should draw circles modelling the node's low range from the centroid. However, due to time constraints, we were unable to test this. Because of time constraints, we were also unable to check that the costs were indeed lessened by choosing either high-level nodes and combining sections to share a higher-level node. Essentially, we overlooked the possible higher cost of the high-band nodes compared to the low-band nodes. We should also examine whether the bandwidth provided by the node decreases with increasing distance.

In terms of the population factor, our model did not take into account the fact that the average time spent per day is not indicative of everyone's time spent. For example, the assumption that those over the age of 60 only use 2 Mbps is not indicative of the mobile device usage of the entire population in that subregion [15]. We also did not consider the number of people in each region who may have technology.

4. Appendix

4.1 Question 1

```
library(scatterplot3d) # This library will allow us to draw 3d plot

speed <- c()
price <- c()
year <- c()
dataset = cbind.data.frame(speed,price,year)
scatterplot3d(speed,price,year)

ls <- function(dataset, par)
        {with(dataset, sum((year-par[1]-par[2]*speed-par[3]*price)^2))}
result <- optim(par=c(0,0,0),ls,data=dataset)

coef <- result$par

plot3d <- scatterplot3d(speed,price,year,angle=235, scale.year=0.7, pch=16, color
="red")
my.lm<- lm(year ~ speed + price,data=dataset)
plot3d$plane3d(my.lm, lty.box = "solid")
summary(my.lm)
```

```
xvar <- c()
yvar <- c()
dataset = cbind.data.frame(yvar,xvar)
lm1 <- lm(yvar~xvar,data=dataset)
summary(lm1)
```

4.2 Question 2

Time Series Plot

```
load 'raw stats analysis - Time Series.csv'
Month = rawStatsAnalysis_TimeSeries.Month;
InternetUsageInc = rawStatsAnalysis_TimeSeries.IncreasingInternetUsageFromBaseline;
tsf = timeseries(InternetUsageInc, 1:69);
tsf.Name = "Internet Usage over Time";
tsf.TimeInfo.Units = "Months";
```

```
tsf.TimeInfo.StartDate = "06, 18";
tsf.TimeInfo.Format = "mm, yy";
tsf.Time = tsf.Time - tsf.Time(1, :);
plot(tsf);
```

Percent Increase Internet Traffic

```
X=[2018:1:2023; 33.45, 41.52, 51.55, 64, 79.46, 98.64];
percentIncrease = [2019:1:2023; (X(2,2)-X(2,1))/X(2,1), (X(2,3)-X(2,2))/X(2,2),
(X(2,4)-X(2,3))/X(2,3), (X(2,5)-X(2,4))/X(2,4), (X(2,6)-X(2,5))/X(2,5)]
```

Month	Internet Usage (GB per month)	Increasing Internet Usage From Baseline		
June 2018	268	1		June 2018 is the
July 2018	268	1		most recent data
August 2018	268	1		found for internet
September 2018	268	1		usage
October 2018	284.08	1.06		
November 2018	300.08	1.119701493		
December 2018	316.08	1.179402985		
January 2019	332.32	1.24		
February 2019	332.32	1.24		
March 2019	332.32	1.24		
April 2019	332.32	1.24		
May 2019	332.32	1.24		
June 2019	332.32	1.24		
July 2019	332.32	1.24		
August 2019	332.32	1.24		
September 2019	332.32	1.24		
October 2019	352.26	1.314402985		
November 2019	372.2	1.38880597		
December 2019	392.14	1.463208955		
January 2020	412.08	1.53761194		
February 2020	412.08	1.53761194		
March 2020	412.08	1.53761194		
April 2020	412.08	1.53761194		
May 2020	412.08	1.53761194		
June 2020	412.08	1.53761194		
July 2020	412.08	1.53761194		
August 2020	412.08	1.53761194		
September 2020	412.08	1.53761194		
October 2020	436.805	1.629869403		
November 2020	461.53	1.722126866		
December 2020	486.255	1.814384328		
January 2021	510.98	1.906641791		
February 2021	510.98	1.906641791		
March 2021	510.98	1.906641791		
April 2021	510.98	1.906641791		
May 2021	510.98	1.906641791		
June 2021	510.98	1.906641791		
July 2021	510.98	1.906641791		
August 2021	510.98	1.906641791		
September 2021	510.98	1.906641791		
October 2021	541.6375	2.021035448		
November 2021	572.295	2.135429104		
December 2021	602.9525	2.249822761		
January 2022	633.61	2.364216418		
February 2022	633.61	2.364216418		
March 2022	633.61	2.364216418		
April 2022	633.61	2.364216418		
May 2022	633.61	2.364216418		
June 2022	633.61	2.364216418		
July 2022	633.61	2.364216418		
August 2022	633.61	2.364216418		
September 2022	633.61	2.364216418		
October 2022	671.6275	2.506072761		
November 2022	709.645	2.647929104		
December 2022	747.6625	2.789785448		
January 2023	785.68	2.931641791		
February 2023	785.68	2.931641791		
March 2023	785.68	2.931641791		
April 2023	785.68	2.931641791		
May 2023	785.68	2.931641791		
June 2023	785.68	2.931641791		
July 2023	785.68	2.931641791		
August 2023	785.68	2.931641791		
September 2023	785.68	2.931641791		
October 2023	832.82	3.107537313		
November 2023	879.96	3.283432836		
December 2023	927.1	3.459328358		
January 2024	974.24	3.635223881		

Weekly usage (hours)

Activity	2-11	12-17	18-34	35-49	50-64	65+
	Age Category					
Watching Traditional Television	13.65	8.77	13.40	26.07	40.83	50.80
TV Connected Game Console	2.92	4.08	3.73	1.62	0.45	0.15
TV Connected Internet Device**	5.55	3.43	5.33	4.95	3.40	2.17
Internet on a Computer (not including video)	0	0	3.70	4.25	4.40	2.77
Video on a Computer	0	0	0.87	0.63	0.50	0.23
Total App/Web on a Smartphone	0	0	24.67	25.02	19.53	13.37
Video Focused App/Web on a Smartphone	0	0	2.50	1.62	0.98	0.58
Steaming Audio on a Smartphone	0	0	0.70	0.57	0.37	0.20
Total App/Web on a Tablet	0	0	4.37	6.03	6.12	7.12
Video Focused App/Web on a Smartphone	0	0	1.05	0.92	0.58	0.60
Steaming Audio on a Smartphone	0	0	0.10	0.15	0.10	0.08

Daily usage (hours)

Activity	2-11	12-17	18-34	35-49	50-64	65+
	Age Category					
TV Connected Game Console	0.4166666667	0.5833333333	0.5333333333	0.230952381	0.06428571429	0.02142857143
TV Connected Internet Device**	0.7928571429	0.4904761905	0.7619047619	0.7071428571	0.4857142857	0.3095238095
Internet on a Computer (not including video)	0	0	0.5285714286	0.6071428571	0.6285714286	0.3952380952
Video on a Computer	0	0	0.1238095238	0.09047619048	0.07142857143	0.03333333333
Total App/Web on a Smartphone	0	0	3.523809524	3.573809524	2.79047619	1.90952381
Video Focused App/Web on a Smartphone	0	0	0.3571428571	0.230952381	0.1404761905	0.08333333333
Steaming Audio on a Smartphone	0	0	0.1	0.08095238095	0.05238095238	0.02857142857
Total App/Web on a Tablet	0	0	0.6238095238	0.8619047619	0.8738095238	1.016666667
Video Focused App/Web on a Smartphone	0	0	0.15	0.130952381	0.08333333333	0.08571428571
Steaming Audio on a Smartphone	0	0	0.01428571429	0.02142857143	0.01428571429	0.0119047619

Assumption: Blue is negligible because we assume that it falls under the 1% category since the time used it so very small --- our model automatically filters for 99%

Usage times of activities

Activity	2-11	12-17	18-64 (H)	18-64 (W)	65+	Daily usage frequencies		
	Age Category							
Watching Traditional Television						ONCE		
TV Connected Game Console	3pm - 9pm	3pm - 9pm	5pm - 12am *	5pm - 12am *	8am - 10pm	ONCE		
TV Connected Internet Device**	3pm - 9pm	3pm - 9pm	5pm - 12am *	5pm - 12am *	8am - 10pm	ONCE	(H) work at home / not working	
Internet on a Computer (not including video)			8am - 10pm	5pm - 12am *	8am - 10pm	MULTIPLE	(W) work away from home	
Video on a Computer			5pm - 12am *	5pm - 12am *	8am - 10pm	ONCE	* denotes the following day	
Total App/Web on a Smartphone			8am - 12am *	5pm - 12am *	8am - 10pm	MULTIPLE		
Video Focused App/Web on a Smartphone				5pm - 12am *	8am - 10pm		for (H)	Anything below 0.16 hours (16 * 0.01) is negligible
Steaming Audio on a Smartphone				5pm - 12am *	8am - 10pm		for (W)	Anything below 0.07 hours (7 * 0.01) is negligible
Total App/Web on a Tablet			8am - 12am *	5pm - 12am *	8am - 10pm	MULTIPLE		
Video Focused App/Web on a Smartphone				5pm - 12am *	8am - 10pm			
Steaming Audio on a Smartphone				5pm - 12am *	8am - 10pm			

4.3 Question 3

```
% For Region A; to test more regions, change information in the variable data
format longG
data = {{690, 1.21, 28.2}, {1422, 0.8, 30.2}, {1303, 0.67, 40.7}, {278, 1.65, 64.3},
{1243, 0.36,37.8}, {1391, 10260,36.9}}; % population, area, and median age data per
subregion given by information sheet
len = length(data);
bandwidth = {};
time = 10260; % Time spent on mobile devices per day for an average person; calculated
from Assumption 1
% Consider population to determine best bandwidth level
M = {};
B = 0; % Set mobile usage on average for the age group
for i = 1:len % Calculate total Mbps used for each subregion
```

```
    if data{i}{3} < 60 % If median age is lower than 60, then set mobile usage on average
to 3
        B = 3;
    else % Otherwise if >= 60 set to 2.
        B = 2;
    end
    M(end+1) = data{i}{1} * time * B;
end
% Assign a bandwidth based on M
for j = 1:len
   if M{j} > 30 && M{j} < 250
       bandwidth{j} = "low";
   elseif M{j} < 900
       bandwidth{j} = "mid";
   elseif M{j} <3000
       bandwidth{j} = "high";
   else
       bandwidth{j} = "more";
   end
end
% Use A = pi*r*r to calculate the smallest range covered by a singular node
low = 3.14*10*10;
mid = 3.14*2*2;
high = 3.14*.5*.5;
% Check to make sure the designated bandwidth can cover the entire
% subsection
for k = 1:len % Compare area to region able to be covered by designated bandwidth
    if bandwidth{k} == "low" % if the area is smaller, than mark that the subsection
needs to be mored
        if data{k}{2} > low
            bandwidth{k} = "more";
        end
    elseif bandwidth{k} == "mid"
        if data{k}{2} > mid
            bandwidth{k} = "more";
        end
    elseif bandwidth{k} == "high"
        if data{k}{2} > high
            bandwidth{k} = "more";
        end
    else
    end
end
for l = 1:len % Print the needed node for each subsection
```

```
    fprintf('Subregion ' + l + ' currently needs ' + bandwidth{l} + 'nodes. \n');
end

disp('Combine regions if possible in order to minimize cost. Then draw the regions on
Geogebra in order to calculate the centroids.');
```

5. Sources

1. https://www.birmingham.ac.uk/schools/ptr/departments/philosophy/news/2020/reglitz-internet-access.aspx
2. Tolbert, Caroline, and Mossberger, Karen, 2015, "U.S. Current Population Survey & American Community Survey Geographic Estimates of Internet Use, 1997-2014", https://doi.org/10.7910/DVN/UKXPZX, Harvard Dataverse, V4
3. https://link.springer.com/article/10.1007/s10922-020-09561-w
4. McConnell, Campbell R., et al. Economics: Principles, Problems, and Policies. McGraw-Hill Education, 2018.
5. https://www.statista.com/chart/23292/monthly-data-usage-america/
6. https://www.sciencedirect.com/science/article/abs/pii/S0167624501000440
7. https://www.wsj.com/articles/americans-working-from-home-face-internet-usage-limits-11603638000
8. https://www.pewresearch.org/internet/2014/03/11/digital-life-in-2025/
9. https://www.statista.com/statistics/216335/data-usage-per-month-in-the-us-by-age/
10. https://decisiondata.org/news/report-the-average-households-internet-data-usage-has-jumped-38x-in-10-years/
11. https://broadbandnow.com/guides/how-much-internet-speed-do-i-need
12. https://www.lenovo.com/us/en/faqs/pc-life-faqs/what-is-mobile-broadband/
13. https://computer.howstuffworks.com/mobile-broadband-service1.htm
14. https://leftronic.com/smartphone-usage-statistics/&sa=D&source=editors&ust=1614626145190000&usg=AOvVaw1MJdIUwvrPAYG9-awf1vNh
15. https://www.nerdwallet.com/article/finance/how-to-decide-what-internet-speed-you-need
16. Defeating the Digital Divide Data, MathWorks Math Modeling Challenge 2021, https://m3challenge.siam.org/node/523.