# MathWorks Math Modeling Challenge 2021

## Adlai E Stevenson High School

Team # 14486 Lincolnshire, Illinois

Coach: Paul Kim

Students: Emily Hu, James Kim, Brandon Lu, Sai Merneedi, Praneet Rathi

## M3 Challenge FINALIST—$6,500 Team Prize

### JUDGES' COMMENTS

*Specifically for Team # 14486 —Submitted at the Close of Triage Judging*:

**COMMENT 1:** Executive summary is fair as some results were provided. Model 1 was reasonable but should have provided R^2 and a discussion. The 10-year predictions are done by graph only and that is not a good way of presenting the results. Model 3 did not address the number of tower.

**COMMENT 2:** Nicely written with clear exposition.

**COMMENT 3:** Very well done and thorough with all questions answered with appropriate models and justified variables. Strengths and weaknesses are just on the points. Nice reading.

# Defeating the Digital Divide

Team 14486

February 28, 2021

# 0    Executive Summary

With the emergence of the COVID-19 pandemic in 2019, the way in which people interact with one another has changed drastically. From schools to businesses, many institutions have learned to take advantage of the internet in order to effectively communicate and carry out daily tasks. However, as the importance of technology has become more apparent, so too has the digital divide. The digital inequality divide is the differences individuals have in the access to the internet — while some are able to take advantage of the growing power of the web, others have fallen behind, including those who cannot afford to access the internet as well as those unfamiliar to technology. The first step to promote greater digital access is to first understand and model how internet and bandwidth access can change over time.

Primarily, in the first part of our paper, we focus on predicting the cost per unit of bandwidth in the US and the UK over the next ten years. By predicting this value, we're able to forecast how consumers and providers will potentially afford bandwidth, and this is the first stage in understanding the digital inequity. We first break down this model into two steps, finding the predicted amount of bandwidth usage for the future and the predicted cost for an average data plan. Due to the varying infrastructure in rural and urban areas, we subcategorized the cost for data plans depending on whether the region was located in a rural or an urban area. Doing so, and combining this result with a weighted average based on population sizes, allows us to get an estimate for the price of bandwidth that takes into account all of America, rather than just cities or urban areas. To calculate the predicted amount of bandwidth used for the future, we utilized Nielsen's law to predict the growth of internet bandwidth usage and extrapolated from existing bandwidth usage data. Putting together the sets of equation yielded a prediction for the average cost per bandwidth for US and UK. The results indicate that the cost of unit bandwidth in the US will be 0.251, 0.080, and 0.026 \$/Mbps in the next 0, 5, and 10 years, respectively. In the UK, the cost per unit bandwidth is projected to be 0.238, 0.107, and 0.048 £/Mbps in the next 0, 5, and 10 years.

In contrast to the national scale in part I, we dive deeper into the individual household's use of internet. In Part II, we are tasked with creating a model that predicts the minimum bandwidth required to cover 90% and 99% of the use. Given the composition of each household in terms of age and income percentile, the model predicts the total bandwidth required to support a household by dividing internet use into two subsections: formal and personal use. Formal use includes school work for children and work for adults done at home since we only take into account the bandwidth need in each household. Furthermore, personal use is categorized based on device used: TV, computer, tablet, and smartphone. Depending on the income percentile and age of each member in a household, we calculate the weekly use of bandwidth in Mb/week. Adding nuance to our model, we factor in the composition of a given household, based on the assumption that other members in a household affect an individual's use of the internet. After taking into account household influences on internet use, we ultimately compute the required bandwidth in Mbps given the composition information of three distinct households. The results show that for the couple in their early 30s, the required bandwidth for 90% and 99% of internet use is 8.76 and 9.64 Mbps, respectively. Similarly, the required bandwidth for household 2 with the retired woman was computed to be 4.10 and 4.51 Mbps. Finally, household 3 with former M3 Challenge participants needed 11.20 and 12.32 Mbps to cover 90% and 99% percent of internet use. The results align with our initial intuition that those who use more internet will require greater bandwidth.

Last, in the third part of our paper, we are tasked with finding a method which will allow us to optimize the placement of nodes/antennas to best meet the bandwidth needs of a specific region. We can calculate the bandwidth needs of a certain region by using the demographics given to us and the model we created in Part II. To create a more accurate result, we calculate the bandwidth needed per square mile using the subregions given, instead of generalizing one statistic for the whole region. After this, we can layer a grid that represents how much area each different type of node/antenna can cover, approximating the area each node covers with a square. By using a smaller square than the actual area of a circle, packing the squares is much more efficient and we can ensure that every spot is covered, at the loss of some efficiency. Using the pre-calculated bandwidth needed per square mile, our model can output exactly where, how many, and what type of nodes are required to meet the minimum bandwidth requirement for everyone.

# 1 Part I: The Cost of Connectivity

In an increasingly digital world, access to reliable internet bandwidth is crucial for functions beyond merely entertainment. From communication, education, to news, the internet has become just as important to our daily lives as food and water are [19]. Given the importance of the internet in our lives, modeling the price of bandwidth over time would allow both customers and providers to predict how internet access could change over time for the future and to create the infrastructure needed to support these changes. Additionally, such a model can help determine how accessible internet will be for the future, especially for those in low-income or particularly remote areas [1].

## 1.1 Defining the Problem

The problem statement asks for a model for the cost per unit of bandwidth in dollars or pounds per Mbps over the next 10 years. To effectively compute the cost per unit of bandwidth, two factors must be computed first—the average amount of bandwidth being used over time, and the average cost of bandwidth over time. Once calculated, these values can be used to calculate the cost per unit bandwidth.

## 1.2 Assumptions

1. *Bandwidth and download speed can be approximated to be the same.*
   Although there are some differences, being that bandwidth is the total available speed, whereas download speed is the bit rate, they both are measurements of internet efficiency [12].

2. *The amount of digital bandwidth usage is predicted to follow an exponential curve for the next 10 years.*
   According to Nielsen's law, bandwidth usage is expected grow at a constant exponential rate, and existing data has supported this observation for the last few decades. Though Nielsen's law is expected to eventually fall off due to limitations in growth, the creation of 5G is predicted to keep Nielsen's law on track for at least the next 10 years [28].

3. *Each city/region follows its own exponential curve for bandwidth usage.*
   Though the growth ratio is constant for all areas, some regions will have head starts due to existing infrastructure. Thus, each city will have the same growth ratio; however, the starting bandwidth usage can vary. Data that shows the change in download speed in the US and UK support the fact that each region resembles a similar exponential growth [23], [24].

4. *We can ignore the cost of one-time installation fees*
   Although upfront costs are a significant factor in the price of broadband, the nature of upfront costs means it isn't logical to incorporate in a cost per unit of bandwidth measurement, since the upfront cost per unit bandwidth would decrease as more bandwidth is used.

5. *The proportion of people living in urban and rural areas will remain the same.*
   Data from the past 10 years shows roughly constant demographics [25].

6. *Consumer buying patterns will remain the same. That is to say, the proportion of a certain region that will buy a certain plan will remain roughly constant.*
   Though on small scales, some individuals will change plans; this shouldn't affect the overall price paid in a region.

7. *Consumers will select the cheapest option for internet given the same internet speed.*
   Assuming that consumers behave optimally most of the time, they will select an internet provider with the lowest cost for their desired internet plan. This assumption is stated in the M3 Spreadsheet on D2 [4].

## 1.3   Parameters

| Symbol | Definition | Units |
|--------|-----------|-------|
| $C_R$ | Average Monthly Cost of Internet for certain region $R$ | $ |
| $BW(t)$ | Download speed (Bandwidth) | Mbps |
| $UC_R(t)$ | Unit Cost of Internet per Bandwidth for a certain region $R$ | $/Mbps |
| $K_{US}$, $K_{UK}$ | A factor that converts Ookla's measurement of internet speed to Akamai's measurement for US and UK respectively | None |

## 1.4   Model Development

To calculate the cost per unit of bandwidth, we first break down our problem into 2 smaller problems. First, we must find how much bandwidth is being used, and then we extrapolate that to model how much total bandwidth will be used over the next 10 years. Second, we must calculate the average monthly cost of bandwidth. Once both of these factors are determined, we can simply divide cost by average bandwidth usage to find the overall cost per unit of bandwidth over time for the next 10 years.

According to assumption 2, we can apply an exponential curve fit to download speed that follows Nielsen's law in order to model the average peak download speed over the next 10 years. Though we could also apply Nielsen's law to average data speed instead of peak data speed, we chose to focus primarily on peak data speed due to the fact that average speed has more variability based on individuals, whereas peak speed is more representative of maximum potential. [11] Using the data set from Akamai and Ookla on average peak download speed, we can use data from 2011 to 2020 in order to create and model a curve fit of average download speed for the next 10 years [4].

However, because Ookla and Akamai use different methodologies to measure download speed, we use a reference year of 2017 to scale Ookla's measurement to the same measurement as Akamai's. Ookla's measurement for peak download speed in 2017 was 70.8 Mbps, whereas Akamai's was 86.5 Mbps. This yields a scaling factor of $K_{US} = \frac{86.5}{70.8} = 1.222$. Repeating the same process for the UK data set yields $K_{UK} = \frac{76.1}{49.2} = 1.5467$ From here, we simply scale the rest of Ookla's data using the same scaling factor to determine our data set.

Using an exponential curve fit yields the following equation for a model seen in Figure 1.1 of Average Peak Download Speed over time where $t$ represents time since 2010.

$$BW_{US}(t) = 17.08 \cdot (1.2566)^t$$

$$BW_{UK}(T) = 21.55 \cdot (1.1742)^t$$

This yields that our exponential growth rate for US to be 1.2566 and an exponential growth rate of 1.1742 for the UK.

Next, in order to find the average monthly cost of internet, we divide the population into two subgroups: urban and rural. According to the Federal Communications Commission, approximately 6 percent, or 19 million people in the US, do not have access to broadband internet with 14.5 million in rural areas [5]. Since, per assumption 5, the population in urban and rural areas will not change dramatically in the next 10 years, we can calculate the number of people in urban and rural areas with and without access to broadband internet.

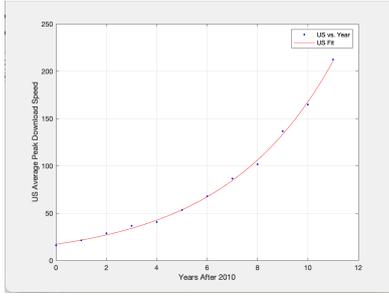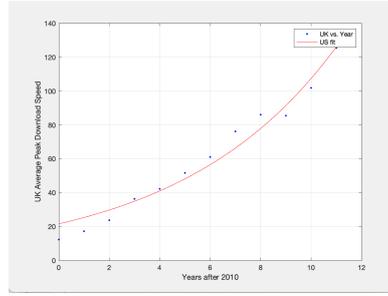| Area | Urban | Rural |
|------|-------|-------|
| With Access to Broadband (millions) | 268.5 | 43.5 |
| Without Access to Broadband (millions) | 4.5 | 14.5 |

Figure 1:
US Average Peak Download
Speed



Figure 2:
UK Average Peak Download
Speed

Figure 1.1: Average Peak Speed (Mbps) vs Years after 2010

Table 1.1: Urban and Rural Population in US with and without Access to Broadband

The cost of internet varies depending on the type. Fiber and cable internet costs an average of $50, while satellite internet is more expensive, costing approximately $100 [22]. In areas without access to broadband, residents must use the satellite option due to the lack of infrastructure for cable and fiber-optic internet. Per assumption 7, residents in areas with broadband access will choose the cheaper option of $50. With this information, we calculate the average monthly cost of internet in the US to be $52.87. Using a similar approach, we calculated the average monthly price of internet in the UK to be approximately £30, which is consistent with the findings of various sources [13], [20].

It is important to note that this breakdown of rural and urban areas is an important breakdown to make. Because the data on bandwidth in various cities from the Open Technology Institute isn't representative of the US as a whole, we must break down all of America instead to have a more complete view of the cost of internet across all of America. However, what we can determine from the Open Technology Institute's data is that the cost families spend on internet plans will stay constant. Reports from 2012, 2014, and 2020 show no significant average change in monthly price, with variations in the same city ranging only by a couple dollars with no apparent pattern.

Thus, it can be justified that the average cost of internet spent in the future will remain roughly constant at $52.87 for the US and £30 for the UK. Then, to find the cost per unit of bandwidth, we can simply take the constant cost we found and divide that by the amount of bandwidth over time.

## 1.5 Results

Combining everything from model development yields the following equations for the average monthly cost of internet for the US and UK as a whole, where $t$ is the number of years since 2021. (Note that the coefficient value is converted by multiplying the original coefficient by $1.2566^{11}$ in order to have $t$ represent time in years since 2021 instead of 2010.)

$$C_{US} = \frac{52.87}{210.7 \cdot (1.2566)^t}$$
$$C_{UK} = \frac{30}{126.06 \cdot (1.1742)^t}$$

Plotting this from $t \in [0, 10]$ (corresponding to the time 2021-2031) yields the plots shown in Figure 1.2. As per assumption 3, we can also use the same model to predict the growth of individual cities as well, using the same growth rate. A city-specific model allows for local governments to determine how their own city
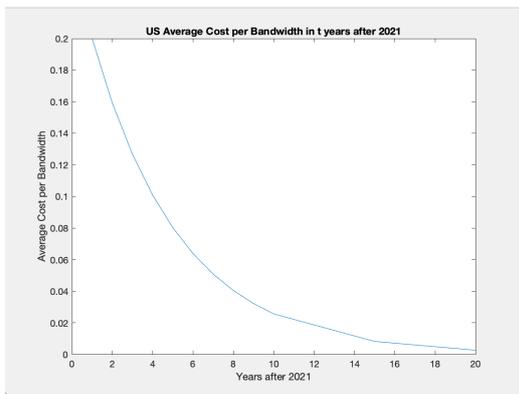
Figure 3:
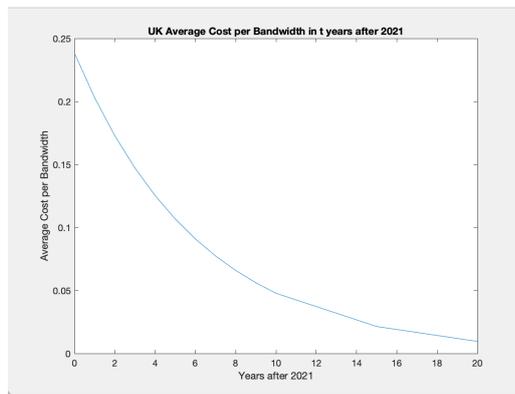US Average Cost per Bandwidth
($/Mbps)



Figure 4:
UK Average Cost per Bandwidth
(£/Mbps)

Figure 1.2: Average Cost per Bandwidth

will progress compared to America as a whole, allowing us to compare and contrast different locations. For example, below is a model for San Francisco, where the initial download speed in 2012 is set to be $41. Using the same growth rate for the US yields the 2021 expected download speed to be $320. Since the average monthly cost of bandwidth is around $60, our final equation modeling the cost per unit bandwidth in San Francisco where $t$ is time in years since 2021 is

$$C_{SF} = \frac{60}{320 \cdot 1.2566^t}$$
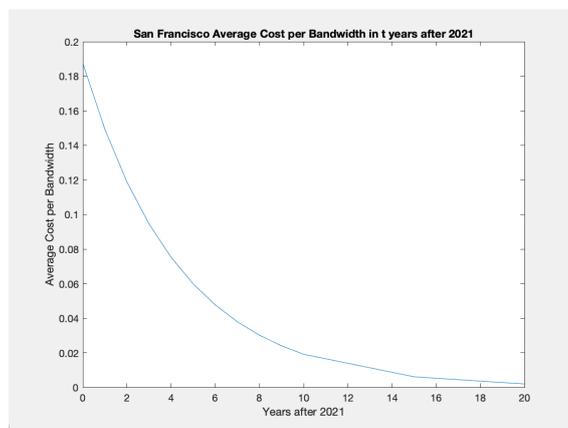
Plotting this from $t \in [0, 10]$



Figure 1.3: San Francisco Average Cost per Bandwidth ($/Mbps)

## 1.6   Sensitivity Analysis

When calculating the average monthly cost in the US, we used the populations in rural and urban areas with and without access to broadband. The cost difference between satellite vs cable and fiber internet option led to varying average monthly cost. The resulting change was produced by modifying the population variables:

| Variable | Percent Change From +5% | Percent Change From -5% |
|---|---|---|
| Urban w/ Access | -0.211 | 0.229 |
| Urban w/o Access | 0.06 | -0.06 |
| Rural w/ Access | -0.0353 | 0.0361 |
| Rural w/o Access | 0.194 | -0.195 |

Table 1.2: Urban and Rural Population in US with and without Access to Broadband

The changes in urban and rural population led to very small changes in the average monthly cost of internet. This indicates that our model is not highly sensitive, which leads to the conclusion that the slight inaccuracies in data will not significantly impact the outcome, validating the results of our model.

## 1.7 Strengths and Weaknesses

The main strengths of the model lies in the fact that it accurately reflects the existing trends from the past years in terms of the growth of download speed. The model is driven by data that shows the average download speed over time during the previous 10+ years and is consistent with trends observed by analysts like Nielsen. Furthermore, the model is efficient in that it can be scaled down to local areas, allowing for local governments to determine and predict how their own region will change over time. By adjusting the growth rate of download speed and the average cost of internet in specific areas, the model can be adapted to fit various scales of interest such as individual cities like San Francisco.

The model has weaknesses in its exponential growth assumption of download speed. Though Nielsen's law is expected to hold for the next 10 years, it is unlikely that this model will hold for time scales larger than that. Like Moore's law, technology growth will eventually be exhausted, and the same is likely to happen with Nielsen's law [28]. In addition, the model does not take into account the changes in the cost of internet, which may lead to inaccuracies in cost per bandwidth in the long term. However, according to the data given, the cost of internet has not significantly changed in the US and the UK in the last 10 years. Again, we chose not to treat average cost as a variable given the short nature of our time frame.

# 2 Part II: Bit by Bit

## 2.1 Defining the Problem

We needed to create a flexible mathematical model to predict a given household's need for the internet. We then needed to apply our model to the example households listed below and determine the minimum amount of required bandwidth that would cover their total internet needs for 90% and 99% of the time.

## 2.2 Assumptions

1. *A household's internet need can be divided into formal use or personal use.*
   Other than for school or work, which is formal use, all other uses of the internet can be categorized as personal use.

2. *All personal use of the internet is on the TV, computer, tablet, or smartphones.*
   A majority of people use those three categories of devices to use the internet. Although a few individuals may use other devices, it's a very small proportion, so it's negligible.

3. *All schools follow the online teaching recommendation from the National Board of Professional Teaching Standards*

The National Board of Professional Teaching Standards is the most prestigious, credible organization, so many schools will want to follow the standard in order to provide the best learning environment for their students. The standard states that elementary students should spend 1.5 hours on online school, middle school students should spend 2.5 hours, and high school students should spend 3.5 hours [17].

4. *Outside of video calling, the bandwidth required for all other school work is negligible.*
   In comparison to video streaming for hours, scrolling through the web is a very small portion of the total broadband usage, which makes it too small to make a difference [7].

5. *Zoom is an accurate estimate of bandwidth needed for all other video conferencing programs.*
   Zoom is the most popular platform currently in the United States, which means a majority of students are using it for their daily school. The receiving and sending HD videos used in Zoom are similar to what other platforms like Google Meets use, so Zoom can act as a good estimate of bandwidth [18].

6. *The probability of an elementary student, middle school, and high school to attend online school are all equivalent.*
   Due to the severity of the pandemic, regardless of age, parents are concerned for the health of their students, so they choose e-learning. Although it seems harder for high school students to learn difficult topics online, it's equally difficult to get a young child to cooperate, making the percentages about equal.

7. *M3 challenge participants are working in a technical field with computers, such as software engineering.*
   M3 challenge is for intelligent STEM students and is based on math and coding, which are crucial topics in computer science and other technical fields with computers. Individuals typically choose to pursue a field that they have previous experience in and are good in.

8. *The school aged grandchildren are between ages 5 and 11.*
   5 is when kindergarten starts and 11 is when children can start taking care of themselves. Typically, if they are above the age of 11, they are in middle school and no longer need a grandparent to watch them [15].

9. *The personal use of internet by a household member can first be influenced by other household members. This potential for influence, as a factor of the original use, decreases with age.*

10. *A college student only has class for 3 hours a day.* On average, a college student needs 15 credit hours in order to graduate in 4 years. 5 hours a week means an average of 3 hours a day [9].

11. *Part-time workers have 20 hour days, and earn half the salary of full time workers.* On average, a college student needs 15 credit hours in order to graduate in 4 years. 5 hours a week means an average of 3 hours a day [9].

## 2.3 Parameters

| Symbol | Definition | Units |
|---|---|---|
| $T_x$ | Weekly internet consumed by member $x$ of a household | \$/week |
| $PF_x$ | Weekly internet consumed for personal use by member $x$ of a household | \$/week |
| $PI_x$ | Weekly internet consumed for personal use by member $x$ of a household only due to $x$'s demographics | \$/week |
| $PO_x$ | Weekly internet consumed for personal use by member $x$ of a household only by influence of other household members | \$/week |
| $F_x$ | Weekly internet consumed for formal purposes—work or school—by member $x$ of a household | MB |
| $P_s$ | Probability of a student going to online school | |
| $h$ | Number of hours for online school per day | hours |
| $I$ | Income Percentile | % |
| $N(I)$ | Number of megabits consumed per week for non-college students as a function of $I$ | MB |
| $C(I)$ | Number of megabits consumed per week for college students as a function of $I$ | MB |

## 2.4 Model Development

To determine an individual's annual use of internet as part of their household, we do so at a weekly level and then scale up to the entire year. In particular, we divide all use of internet of an individual in a household into two primary categories: for formal use and for personal use, per assumption 1. This can be represented as follows.

$$T_x = F_x + P_x$$

Per the examples given in the problem, each individual of the household is defined by their age and, where it is applicable, education/working status. However, there must be some other factor that varies from household to household given the same composition which leads to the 90% and the 99% thresholds of the bandwidth requirements for any of these households. We reason that income is this factor, given the notable difference uses of internet dependent on income, such that we determine the average bandwidth consumed within a household over the course of a year for a given composition, across all incomes. Then, whatever bandwidth satisfies the average need for 90% of household incomes, again with the composition constant, becomes the 90% threshold or, in the words of the prompt, covers total internet needs 90% of the time.

**Formal Use**

The two formal uses we considered are internet used for school and the internet used for work.

For school, only individuals who choose e-learning will need to use the household internet for school since the school will be the one providing the internet if the student is at school. Analyzing previous data, we notice that a household's income level affects the probability of a student learning at home. Furthermore, among all students who do e-learning, different ages have different internet needs due to the duration of school. Per assumption 4, the main source of internet need is caused by video calling, so that is what we used to calculate the internet need per person per week.

Taking these two variables of income and also age of student, we are able to develop a model that relates the household income with the weekly gigabit usage for each level of school. We grouped elementary, middle,

and high school into one model and college students in the other one. We specifically chose these groupings since, before college, the school system is very similar.

Only focusing on elementary, middle, and high schools, we were able to determine the percentage of students who chose online learning given their household income where low income is the 10% percentile of household incomes, middle income is 50% percentile, and high income is 90% percentile.

| Probability of non-college students having online school based on household income (Table 2.1) [3] | | | |
|---|---|---|---|
| | Low income | Middle Income | High Income |
| Probability of having online school | 54.2% | 58.9% | 64.9% |

Per assumption 5 and 3, we can multiply the probability of a student of a certain household income to attend school by the total hours a student spends on online school and the total bandwidth Zoom needs to find the total number of megabits needed for one day of school. The bandwidth required for Zoom is 4.8 Mbps, but since we want to find the number of megabits in an hour, we multiplied the rate by $60^2$, which is the number of seconds in an hour.

$$\text{Total number of megabits} = P_s \cdot 4.8 \cdot 60.^2 \cdot h$$

After calculating, we were able to attain the following table of values that related the household income to the number of megabits used per week by online school.

| The number of megabits consumed daily based on household income (Table 2.2) | | | |
|---|---|---|---|
| | Low income | Middle Income | High Income |
| Elementary School | 14,048 | 15,290 | 16,814 |
| Middle School | 23,414 | 25,440 | 28,020 |
| High School | 32,780 | 35,616 | 39,230 |

Taking these values in the table and plotting it on a graph, we were able to get the equation of the curves, which can be used to calculate the number of megabits used per week for online school given the school type and household income level.
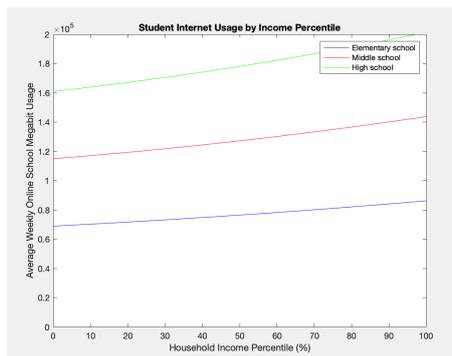


Figure 5: Student Internet Use by Income Percentile (MB)

$$N_{\text{elementary school}} = 8.812 \cdot 10^{-5} \cdot I^2 + 0.02576 \cdot I + 13.78$$
$$N_{\text{middle school}} = 0.0001731 \cdot I^2 + 0.04026 \cdot I + 22.99, \quad N_{\text{high school}} = 0.0002431 \cdot I^2 + 0.05631 \cdot I + 32.19$$

Then for college we take a similar approach, where we find the probability that the college student is taking online courses for each income percentile first in order to find the number of megabits they consume a week. Using the data found, we were able to create the following table of values.

| Probability of college students having online school based on household income (Table 2.3) [8] | | | |
|---|---|---|---|
| | Low income | Middle Income | High Income |
| Probability of having online college | 13% | 12% | 7% |

Using the same value of 4.8 mbps for the bandwidth Zoom required and 3 hours per day due to assumption 10, we can calculate the total megabits of internet needed per day. We multiplied together the 4.8 mbps, the $60 \cdot 60$ from the 3600 seconds in an hour, and the 3 hours of college per day to find the total megabits.

| Internet use in Megabits by college Students per day(Table 2.4) | | | |
|---|---|---|---|
| | Low income | Middle Income | High Income |
| Total Number of Megabits per day | 6,739 | 6,220 | 3,628 |
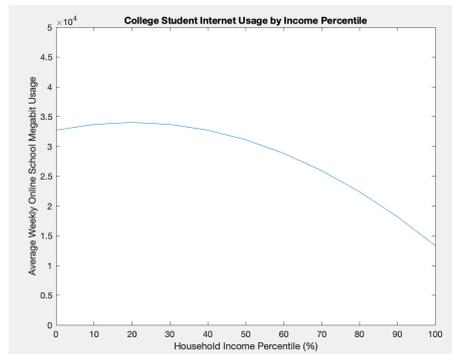


Figure 6: College Student Internet Use by Income Percentile (MB)

$$C = -0.000647 \cdot I^2 + 0.02589 \cdot I + 6.545$$

The other part of formal use is internet use for work. This depends on the nature of this work, whether this is physical work, mixed work, or completely online work. Given the large range of function inputs and possible range, we therefore solve for this in each the specific use cases.

**Personal Use**

Personal use per assumption 1 is defined to be any use of the internet besides official school or work. In particular, per assumption 2, we seek to model personal use of television, computers, tablets, and smartphones. We determine this use by an individual's age and income, with age being given in the problem's use cases, and income being what determines the confidence thresholds.

However, per assumption 9, personal use is not only defined by individual age and household income, which gives the standalone $PI_x$, but also defined by the other household members' uses, $PO_x$.

$$PF_x = PI_x + PO_x$$

We say that there exists some pairwise influence of one household member's personal use of the internet on another household member's personal use of the internet. This follows from siblings on average taking more interest and ultimately using the internet in more ways similar to each other, children having uses of the internet dependent on their parents uses, and many more pairwise relations. We therefore use the pairwise $o_{x1x2}$ as follows, where 1 is the member we're solving for and $2 \cdots n$ are other members of the household.

$$PO_1 = PI_1 + PI_2 o_{12} + \cdots + PI_n o_{1n}$$

To convey this at a household rather than member of household level, we can use matrix multiplication.

$$
\begin{bmatrix} PI_1 & PI_2 & \cdots & PI_n \end{bmatrix}
\begin{bmatrix}
1 & o_{12} & \cdots & o_{1n} \\
o_{21} & 1 & \cdots & o_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
o_{n1} & o_{n2} & \cdots & 1
\end{bmatrix}
$$

We first look at $PI_x$ for any household member $x$ as defined by their age and income. Per assumption 2, this can be rewritten as follows.

$$
PI_x = PI_{tv}x + PI_{comp}x + PI_{tablet}x + PI_{phone}x
$$

The subscript label of each defines what the internet was used for, such that for example $PI_{tv}x$ represents the weekly personal use of TV in the household by household member $x$. We first look at how $PI_{tv}x$ can be calculated and then use the same methodology to determine $PI_{comp}x$, $PI_{tablet}x$, and $PI_{phone}x$. The weekly TV use of internet, from the data given, can be determined by the total time spent by the member using the internet for TV, $t_{tv}x$, multiplied by the expected value of the bandwidth for using TV, $b_{tv}x$. Again because this is being taken across families to determine the annual averages, we multiply by bandwidth by time to get aggregate internet usage for a year and then divide by time in a year to get average bandwidth at any time. However, these families vary by income, and there are clear trends relating the income of families and their use of particular internet devices like TV [4].

We therefore establish an income multiplier $m_{tv}(I)$ that is a function of $I$, the household income. What this means is given the factor of age constant, being a member of households with different total incomes will lead to different television use. To determine this multiplier, we fit a curve relating the household income and average television use, using the provided M3 data [4].
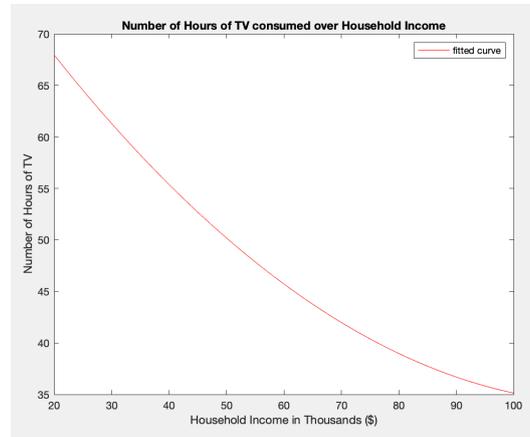


Relation between household income and hours of television used

To get the actual multiplier, however, we must divide the output of this function by the average output of the function. Performing this same process for each of TV, computers, tablets, and phone we get the following expression:

$$
M_{\text{TV}} = \frac{0.003637I^2 - 0.8467I + 83.43}{51.15}, \quad M_{\text{comp}} = \frac{25.53I^{-0.2559}}{9.62}
$$

$$
M_{\text{tablet}} = \frac{-0.006318I + 7.185}{6.84}, \quad M_{\text{phone}} = \frac{7.643 \cdot 10^{-5}I^2 - 0.03214I + 13.55}{11.92}
$$

We thus get $t_{tv}x \cdot b_{tv}x \cdot m_{tv}(I)$ to be the expression for $PI_x$, where the $x$ is the member's age, and we know $I$ is the income of the household the member belongs in. We must determine what the remaining expression $PI_2 o_{12} + \cdots + PI_n o_{1n}$ for the expression of individual personal use is, or in

other words the effect other household members have on a given member. Following assumption 9, we propose that $PI_2o_{12} + \cdots + PI_no_{1n}$ depends entirely on the age of person 1. We fit the following curve using data relating age and capacity to obtain digital literacy, reasoning that this capacity is the same as the capacity to have internet usage affected [14].
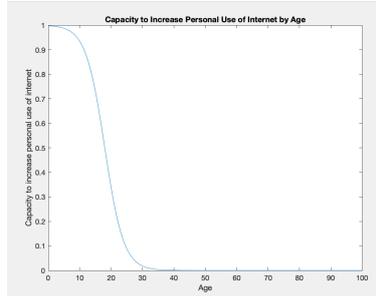


Relation between age and capacity to have internet usage affected

Expressing this relation as $c(1)$ where the age of person 1 is input to determine capacity, we determine that $PI_2o_{12} + \cdots + PI_no_{1n} = c(1)$.

Putting it together to get $PF_x$, we combine this with $F_x$ to get $T_x$.

## 2.5 Model Enactment

**Household 1**

We first looked at the adult who's a teacher and created a function that relates his/her use of the internet as a multiplier of the student's use of the internet to find formal use of the internet. The child is 3 and therefore has no formal internet use. For the personal use of each of the adults and the child, we took into account the household income, made only by the teacher and calculated using the age parameter of 32.5, and the individual ages. Using previous data on data consumption for each of these, we created a function based on household salary [16][4] [10] [21].
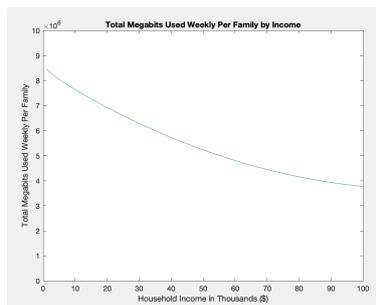


Relation between Megabits consumed and income for the first household
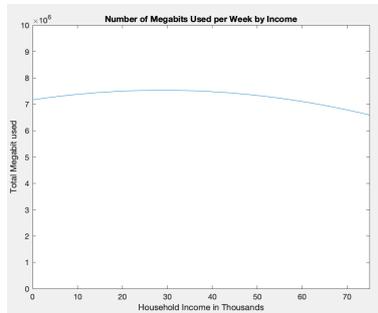
**Household 2**

We then looked at the woman and used data for what average retirement income given her age is, to determine her personal use. Using the age per assumption 8, and assuming they come to her only for school, we determined their formal use for the time they are there [27] [6].
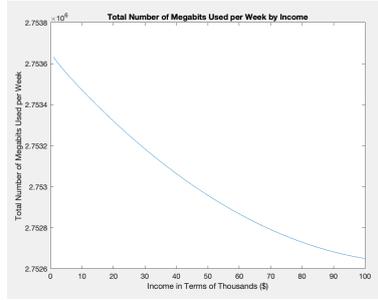
Relation between Megabits consumed and income for the second household

**Household 3**

Since the three individuals are in college, we calculate the bandwidth for each assuming they all use bandwidth dependent on what they're making. We used ages of 20 to get formal use of education and formal use of work, the former calculated with our education equation and the latter using online data [26].



Relation between Megabits consumed and income for the third household

Finding the distribution of internet consumed over the course of a week by income, we found 90% and 99% over possible income values, and divided by 3600, 7,and 24 to find the bandwidth needed per second. The first household require 8.76 and 9.64 Mbps Similarly, the required bandwidth for household 2 with the retired woman was computed to be 4.10 and 4.51 Mbps. Finally, household 3 with former M3 Challenge participants needed 11.20 and 12.32 Mbps to cover 90% and 99% percent of internet use.

| Variable | Household 1 | Household 2 | Household 3 |
|----------|-------------|-------------|-------------|
| 90%      | 8.76        | 4.10        | 11.20       |
| 99%      | 9.64        | 4.51        | 12.32       |

Table 2.5: 90% and 99% bandwidth required (Mbps)

## 2.6   Strengths and Weaknesses

The model is effective in that it reflects the current state of the country well. By taking into account the level of internet use due to remote working and education environments as a result of the COVID-19 pandemic, the model accurately predicts the required bandwidth percentiles based on internet use at any given time. Also, by understanding the relationship between members of a household, the model generates accurate outcomes as it relates to the composition of a household, which was the most significant factor given in the prompt.

However, the model has some weaknesses in that it takes the average required bandwidth per second rather than modeling its change as a function of time. This lead to the calculation of 90% and 99% percentile to

be based on income percentile. However, our approach to the problem reflects a different interpretation of the task, which lead to unique results and a wide range of factors that determine the required bandwidth.

# 3 Part III: Mobilizing Mobile

## 3.1 Defining the Problem

As they are the easiest form of internet access, mobile phones are important factors in solving the digital inequity of our society. In order to do so, 4G and 5G connections are established through cell towers, or nodes, that provide connection within a certain radius [29]. In Part III, we are tasked to find an optimal strategy for placing the cellular nodes in various regions.

We first define optimal strategy as building the least number of cell towers, or nodes while still satisfying the bandwidth need of the region. In order to find this, our model divides the region into smaller subdivisions to calculate the minimum number of nodes required to supply the bandwidth need of each division.

## 3.2 Assumptions

1. *Bandwidth need is evenly distributed across each subregion.*
   Bandwidth need density is defined as the total bandwidth needed per unit area in a given subregion. Since no additional information is given about the distribution of population within each subregion, we can simplify the calculations by assuming an even distribution of bandwidth need across each subregion based on factors like median age and income. This bandwidth need density can be calculated with the model from part II.

2. *The range of each cell tower is defined to be a square area with fixed dimensions.*
   For simplicity, we assume that each node will be able to provide internet connectivity for a certain region with dimensions $a \times a$, where $a$ is the range in miles. There are three types of cell towers (low, mid, and high band) and the corresponding ranges are obtained from the M3 Data source, D9 [2]. In doing so, packing the squares to determine whether each space is covered is a lot more efficient, and by under-approximating the area, we ensure that there are no empty spaces with no coverage.

3. *Cell tower bandwidths can be added together for additional strength.*
   For example, having a cell tower with bandwidth of 10 Mbps and another tower with bandwidth of 5 Mbps means that the overlapping region will have bandwidth of 15 Mbps.

## 3.3 Parameters

| Symbol | Definition | Units |
|---|---|---|
| $R_L, R_M, R_H$ | Range of Low, Mid, and High Band | miles |
| $DS_L, DS_M, DS_H$ | Average Download Speed of Low, Mid, and High Band | Mbps |
| $BD_S$ | Bandwidth Need Density for Subregion $S$ | Mbps/miles$^2$ |
| $B_G$ | Bandwidth Need for Grid $G$ | Mbps |
| $A_G$ | Area of Grid $G$ | miles$^2$ |
| $N_G$ | Number of Cellular Nodes Needed for Grid $G$ | nodes |
| $N_T$ | Total Number of Cellular Nodes Needed for Region | nodes |

## 3.4　Model Development

The optimal strategy is defined to be the placement of cellular nodes so that the number of nodes is minimized while satisfying the bandwidth need of the entire population. Depending on the bandwidth need, which can be determined in part II based on demographics like median age and income, we calculate the bandwidth need density of a certain subregion S, $BD_S$, by dividing the total bandwidth need in a subregion by its area. Generally, in areas with high bandwidth need density, a higher band cell tower would be optimal. Conversely, in areas with low bandwidth need density, a low band tower would be ideal. From the Mobile broadband frequency band characteristics sheet, we calculate the average download speed and range of low, mid, and high band cell nodes [2].

| Band | Average Download Speed | Average Range |
|------|------------------------|---------------|
| Low  | $DS_L = 140$           | $R_L = 15$    |
| Mid  | $DS_M = 500$           | $R_M = 3$     |
| High | $DS_H = 2000$          | $R_H = 0.75$  |

Since the range of the high band cell tower is the smallest, we start the model by dividing the entire region into square grids with dimensions $R_H \times R_H$. For example, Figure 3.1 illustrates Region B divided into 0.75 mile $\times$ 0.75 mile square grids, which were scaled using the pixel count and given area of the region.
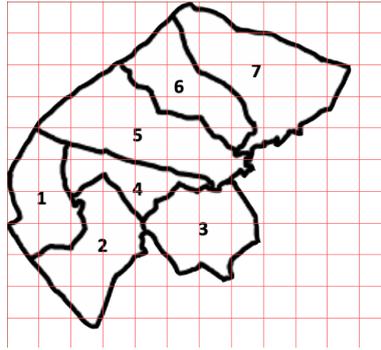


Figure 3.1: Region B divided into .75 mile sided squares, each representing a high band node's coverage

We then calculate the total bandwidth need of each grid, $B_G$, by using the maximum bandwidth need density of the subregions that lie inside the grid, multiplied by the area of the grid, $A_G$, effectively assuming that the maximum bandwidth is needed for the whole square. For example, if a grid contains portions of 3 subregions with bandwidth need densities of 10, 30, and 50 $Mbps/mile^2$, the bandwidth need of the grid will be determined using the maximum, 50 $Mbps/mile^2$, multiplied by $A_G$. In proper notation, this is represented by the following equation:

$$B_G = A_G \cdot \max(BD_{S1}, BD_{S2}, ..., BD_{S_n})$$

Since one high band cell tower can provide an average of $D_H = 2000$ Mbps, the number of high band nodes is calculated with the following equation:

$$N_G(\text{high}) = \left\lfloor \frac{B_G}{DS_H} \right\rfloor$$

In order to reduce inefficiency caused by providing an over abundance of bandwidth, we take the floor of $BD_H/D_H$, which leaves some bandwidth to be covered by lower band cell towers. Per assumption 3, since the bandwidth of multiple towers can be added together in a given grid, the total bandwidth of any grid can be composed of multiple types of cell towers. In order to do so, we subtract the bandwidth added by all the high band towers from the square that the high band tower is in. For example, if the bandwidth required in a square is 4500, after adding 2 high band cell towers, the remaining bandwidth need is now

500. Hence, with the remaining bandwidth needed, we move on to a lower band cell tower and change the grid size to its corresponding range.

We then repeat this process for the mid band node and low band node, with the exception that the low band node should be a ceiling function since we want to make sure the bandwidth requirement is met. For example, in Figure 3.2, we have region B divided into 3 mile sided squares, representative of the coverage of a mid band node, where we would apply the same process as what we did with the high band nodes to determine whether nodes are needed in each square.
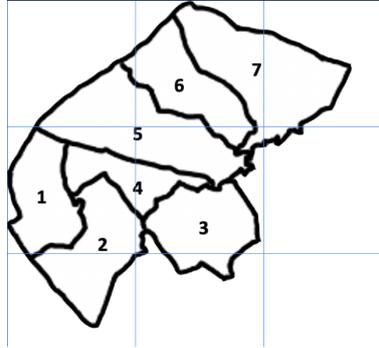


Figure 3.2: Region B divided into 3 mile sided squares, each representing a mid band node's coverage

Formally, the number of low, mid, and high band cell nodes required for grid $G$ is calculated from the following:

$$N_G(\text{high}) = \left\lfloor \frac{B_G}{DS_H} \right\rfloor, \; N_G(\text{mid}) = \left\lfloor \frac{(B_G - N_G(\text{high} \cdot DS_H))}{DS_M} \right\rfloor, \; N_G(\text{low}) = \left\lceil \frac{B_G - N_G(\text{high} \cdot DS_H) - N_G(\text{mid}) \cdot DS_M)}{DS_L} \right\rceil$$

Ultimately, we sum the total number of low, mid, and high band nodes to calculate the total number of cell towers needed to satisfy the bandwidth demand for a given region.

$$N_T = \sum_{G=1}^{h} N_i(\text{high}) + \sum_{G=1}^{m} N_i(\text{mid}) + \sum_{G=1}^{l} N_i(\text{low})$$

After applying the algorithm, we will end up with a map that depicts exactly how many, where, and type of nodes/cell towers will be needed to meet the bandwidth requirement of everyone in the region. This solution is close to optimal since the areas with the highest bandwidth need density are the first to have cell towers installed, and it guarantees that all bandwidth demand is met.
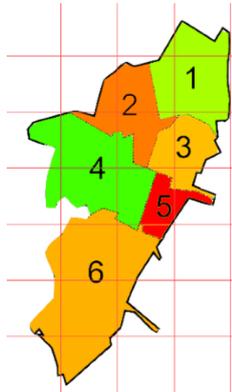
## 3.5 Results

First, applying our model from part II to region A, will yield the following bandwidth demands from each subregion.
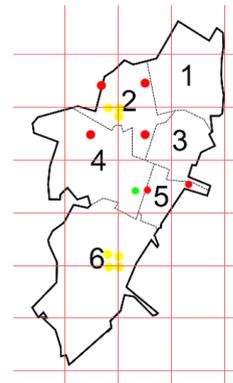
| subregion | Bandwidth Demand Per Area |
|---|---|
| 1 | 814.0 |
| 2 | 2540.2 |
| 3 | 1867.7 |
| 4 | 242.2 |
| 5 | 3517.3 |
| 6 | 1534.5 |

Figure 3.3 depicts a visual representation for the data above, where we colored the map of region A, using darker red colors to represent higher bandwidth demand density and lighter green colors to represent lower bandwidth demand density. Additionally, Figure 3.3 also shows the result after applying the algorithm to

the region. The resultant locations for high, mid, and low band nodes are depicted with red, yellow, and green circles, respectively. Note that nodes are placed in the center of their respective grid spaces, such that they can reach everything inside the respective grid space, unless the center is outside of the region, in which case the node is placed on the border.
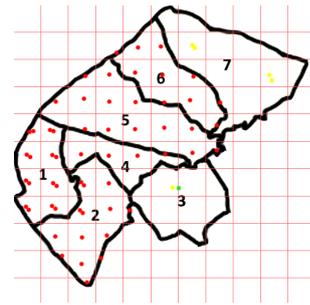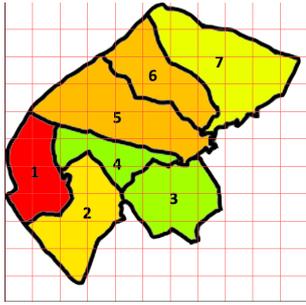


Colored region A map depicting bandwidth demand density



Location of nodes and the type for region A
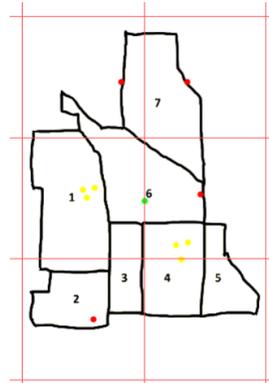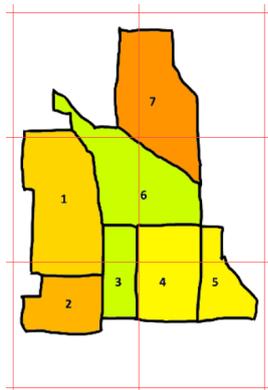
Figure 3.3: Region A results

Applying the same algorithm to Region B and Region C yields the figures seen in figure 3.4 and 3.5



Colored Region B map depicting bandwidth demand density



Location of nodes and the type for region B

Figure 3.4: Region B results



Colored Region C map depicting bandwidth demand density



Location of nodes and the type for region C

Figure 3.5: Region C results

## 3.6   Strengths & Weaknesses

The strength of this model is that the process is one that is structured and algorithmic. This means that our model can be repeated for various other shapes and produce consistent results. Furthermore, because our model is an algorithm, much of this work can be computed automatically, rather than by hand or with other methods, drastically lowering computation time. Furthermore, because every assumption we made was under-approximation of the capabilities of our nodes, we are sure that every location in every region will have the sufficient bandwidth required.

However, because we did make under-approximations when estimating the area each node covered as well as how much bandwidth is required in each grid square, our model will produce a moderately inefficient output. This means that there is a chance that the model may over count the number of nodes needed. This, however, is expected because the nature of packing problem are heuristic and classified as NP-hard, meaning that the problem cannot be solved in polynomial time, and hence, there is no algorithmic way to calculate the optimal way for every region.

# References

[1] Digital inequality and low-income households, 2016. URL: `https://www.huduser.gov/portal/periodicals/em/fall16/highlight2.html`.

[2] Venture Beat. The definitive guide to 5G low, mid, and high band speeds, 2019. URL: `https://venturebeat.com/2019/12/10/the-definitive-guide-to-5g-low-mid-and-high-band-speeds/`.

[3] Census.gov. Week 16 household pulse survey: September 30–October 12, 2020. URL: `https://www.census.gov/data/tables/2020/demo/hhp/hhp16.html#tables`.

[4] MathWorks Math Modeling Challenge. Cost for bandwidth (US, wired), 2021. URL: `https://m3challenge.siam.org/node/523`.

[5] Federal Communications Commission. Eighth broadband progress report. URL: `https://www.fcc.gov/reports-research/reports/broadband-progress-reports/eighth-broadband-progress-report`.

[6] Amin Dabit. What is the average retirement income and how do you compare?, 2020. URL: `https://www.personalcapital.com/blog/retirement-planning/average-retirement-income/`.

[7] Doug Dawson. Bandwidth needed to work from home, 2020. URL: `https://www.circleid.com/posts/20200803-bandwidth-needed-to-work-from-home/`.

[8] Erin Duffin. Share of students studying online in the united states in 2020, by annual household income in U.S. dollars and education level, 2020. URL: `https://www.statista.com/statistics/956154/share-students-studying-online-income-education-level/`.

[9] Abigail Endsley. How many credits do I need for a bachelor's degree?, 2017. URL: `https://pearsonaccelerated.com/blog/how-many-credits-do-i-need-for-a-bachelors-degree`.

[10] Fairygodboss. How many hours is part time? URL: `https://fairygodboss.com/career-topics/how-many-hours-is-part-time#:~:text=A%20part%2Dtime%20employee%20might,a%20full%2Dtime%20job`.

[11] Fastmetrics. Average peak & median internet speeds USA, 2020. URL: `https://www.fastmetrics.com/internet-connection-speed-map-usa.php`.

[12] Greg Ferro. Basics: The difference between bandwidth and speed, 2015. URL: `https://etherealmind.com/basics-difference-bandwidth-speed/`.

[13] Glide. Average internet bill in the UK.
URL: `https://glide.co.uk/guides/average-internet-bill-uk/`.

[14] Kids Guitar. What is the best age to start learning guitar? URL: `https://kidsguitar.info/best-age-to-start-learning-guitar/`.

[15] Sarah Henry. Developmental milestone: Self-care (ages 6 to 8), 2019. URL: `https://www.babycenter.com/child/development/developmental-milestone-self-care-ages-6-to-8_67734`.

[16] Nielsen Holdings. The total audience report, 2015. URL: `https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/total-audience-report-q3-2015.pdf`.

[17] Hannah Hudson. Kids (and teachers) don't need to spend 8 hours a day on school work right now, 2020. URL: `https://www.weareteachers.com/virtual-learning-schedule/`.

[18] Rani Molla. The pandemic was great for Zoom. what happens when there's a vaccine?, 2020. URL: `https://www.vox.com/recode/21726260/zoom-microsoft-teams-video-conferencing-post-pandemic-coronavirus`.

[19] Diane Myers. Top 10 uses of the internet, 1995. URL: `https://www.edn.com/top-10-uses-of-the-internet/`.

[20] UK Office of Communications. Pricing trends for communications services in the UK, 2020. URL: `https://www.ofcom.org.uk/__data/assets/pdf_file/0022/189112/pricing-trends-communication-services-report.pdf`.

[21] National Center of Education Statistics. The patterns of teacher compensation, 1996. URL: `https://nces.ed.gov/pubs/web/95829.asp`.

[22] Dave Schafer. How much does satellite internet cost?, 2019. URL: `https://www.satelliteinternet.com/resources/how-much-does-satellite-internet-cost/`.

[23] Statista. Average internet connection speed in the United States from 2007 to 2017 (in mbps), by quarter, 2017. URL: `https://www.statista.com/statistics/616210/average-internet-connection-speed-in-the-us/`.

[24] Statista. Internet connection speed in the United Kingdom (UK) 2007-2017, 2017. URL: `https://www.statista.com/statistics/475965/internet-connection-speed-in-the-uk/`.

[25] Statista. Size of urban and rural population of the U.S. 1960-2020, 2020. URL: `https://www.statista.com/statistics/985183/size-urban-rural-population-us/#:~:text=Size%20of%20urban%20and%20rural%20population%20of%20the%20U.S.%201960%2D2020&text=In%202020%2C%20there%20were%20approximately,people%20living%20in%20urban%20areas`.

[26] Burning Glass Technologies. Real-time insight into the market for entry-level STEM jobs, 2014. URL: `https://www.burning-glass.com/wp-content/uploads/Real-Time-Insight-Into-The-Market-For-Entry-Level-STEM-Jobs.pdf`.

[27] National Telecommunications and Information Administration. Chapter 5: The digital generation: How young people have embraced computers and the internet. URL: `https://www.ntia.doc.gov/legacy/ntiahome/dn/html/Chapter5.htm#:~:text=%EF%BF%BD%20Within%20the%20same%20age,(Figure%205%2D2)`.

[28] John Ulm. Is Nielsen ready to retire?, 2014. URL: `https://www.commscope.com/globalassets/digizuite/1642-arris-is-nielsen-ready-to-retire-wp.pdf`.

[29] Verizon. 5G nodes: what they are and how they work, 2020. URL: `https://www.verizon.com/about/news/towers-what-they-are-how-they-work`.

# 4   Code Appendix in Matlab

Part I: Derive Exponential Graph of Download Speed as a function of time

```matlab
\haven % Fit: 'US Fit'.
[xData, yData] = prepareCurveData( Year, US );

% Set up fittype and options.
ft = fittype( 'exp1' );
opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
opts.Display = 'Off';
opts.StartPoint = [17.2043880140776 0.227554602424087];

% Fit model to data.
[fitresult, gof] = fit( xData, yData, ft, opts );

% Plot fit with data.
figure( 'Name', 'untitled fit 1' );
h = plot( fitresult, xData, yData );
legend( h, 'US vs. Year', 'US Fit', 'Location', 'NorthEast', 'Interpreter', 'none' );
% Label axes
xlabel( 'Years After 2010', 'Interpreter', 'none' );
ylabel( 'US Average Peak Download Speed', 'Interpreter', 'none' );
grid on

%% Fit: 'UK fit'.
[xData, yData] = prepareCurveData( Year, UK );

% Set up fittype and options.
ft = fittype( 'exp1' );
opts = fitoptions( 'Method', 'NonlinearLeastSquares' );
opts.Display = 'Off';
opts.StartPoint = [18.659763439111 0.178741691340209];

% Fit model to data.
[fitresult, gof] = fit( xData, yData, ft, opts );

% Plot fit with data.
figure( 'Name', 'US fit' );
h = plot( fitresult, xData, yData );
legend( h, 'UK vs. Year', 'US fit', 'Location', 'NorthEast', 'Interpreter', 'none' );
% Label axes
xlabel( 'Years after 2010', 'Interpreter', 'none' );
ylabel( 'UK Average Peak Download Speed', 'Interpreter', 'none' );
grid on
```

Part I: Plot Cost per Bandwidth in US, UK, and San Francisco

```matlab
year = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20]'; % t years after 2021
plot(year, 52.87 ./(210.6*exp(0.2284*year))) %plot cost per bandwidth in us
title("US Average Cost per Bandwidth in t years after 2021")
xlabel('Years after 2021')
ylabel('Average Cost per Bandwidth')
```

```
    year = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20]'; % t years after 2021
    plot(year, 30 ./(126*exp(0.1606*year))) %plot cost per bandwidth in uk
    title("UK Average Cost per Bandwidth in t years after 2021")
    xlabel('Years after 2021')
    ylabel('Average Cost per Bandwidth')

    year = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20]'; % t years after 2021
    plot(year, 60 ./(320*exp(0.2284*year))) %plot cost per bandwidth in San Francisco
    title("San Francisco Average Cost per Bandwidth in t years after 2021")
    xlabel('Years after 2021')
    ylabel('Average Cost per Bandwidth')
```

Part II

```
i= 0:10:100
i=0:1:75
broadband=180000+3*3600*(-0.000647*(0.7931*i-2.828).^2+0.02589*(0.7931*i-2.828)+6.545)*100
plot(i, broadband)
xlim([0 75])
ylim([0 10000000])
xlabel('Household Income in Thousands')
ylabel('Total Megabit used')
title('Number of Megabits Used per Week by Income')

incomelevel= 0:10:100
elem= [14.048 15.29 16.814]
mid= [23.414 25.44 28.02]
high= [32.78 35.616 39.23]
plot(incomelevel, elem)
xlim([0 100])
ylim([0 100])
hold on
plot(incomelevel, mid)
plot(incomelevel, high)

legend('Elementary school', 'Middle school', 'High school')

x= 0:10:100
x=83.41
elem= (8.812*10.^(-5)*x.^2+0.02576*x+13.78)*5000
mid= (0.0001731*x.^2+0.04026*x+22.99)*5000
high= (0.0002431*x.^2+0.05631*x+32.19)*5000
avg=(elem+mid+high)./3
plot(x, elem, 'b-')
xlim([0 100])
ylim([0 200000])
hold on
plot(incomelevel, mid,'r-')
plot(incomelevel, high,'g-')
title('Student Internet Usage by Income Percentile')
xlabel('Household Income Percentile (%)')
ylabel('Average Weekly Online School Megabit Usage')
L(1) = plot(nan, nan, 'b-');
L(2) = plot(nan, nan, 'r-');
L(3)=plot(nan, nan, 'g-')
legend(L, {'Elementary school', 'Middle school', 'High school'})
```

```
income= [20 37.5 62.5 125]
TV= [68.32 56 45.27 35.02]
Comp= [12.01 9.71 9.12 7.85]
phone=[13.14 12.05 12.05 10.43]
tablet=[6.84 7.12 6.97 6.42]
i=0:1:150
eq=3600*(22.46*(19.3)*(TVfit(i)./51.15)...
+2*((TVfit(i)./51.15)*19.3*22.46+4.14*Compfit(i)./9.67*4.57+2.07*phonefit(i)./11.92*27.87...
+tabletfit(i)./6.84*2.07*5.52))
plot(i, eq)
xlim([0 100])
ylim([0 10.^7])
xlabel('Household Income in Thousands ($)')
ylabel('Total Megabits Used Weekly Per Family')
title('Total Megabits Used Weekly Per Family by Income')


x= 0:1:100
y=1-(1./(1+exp(-1./3*(x-18))))
plot(x, y)
xlabel('Age')
ylabel('Capacity to increase personal use of internet')
title('Capacity to Increase Personal Use of Internet by Age')


plot(TVfit)
xlabel('Household Income in Thousands ($)')
ylabel('Number of Hours of TV')
title('Number of Hours of TV consumed over Household Income')


i=0:1:100
total=3600*((8.812*10.^(-5)*50.^2+0.02576*50+13.78)*50)...
+((TVfit(i)./51.15)*19.3*53.12+4.14*Compfit(i)./9.67*3+2.07*phonefit(i)./11.92*14.15+tabletfit(i)./6.84*2.07*7.8)
plot(i, total)
ylabel('Total Number of Megabits Used per Week')
xlabel('Income in Terms of Thousands ($)')
title('Total Number of Megabits Used per Week by Income')
```