1 Executive Summary

The COVID-19 pandemic engendered many changes to society. One of the largest shifts was the unprecedented increase in remote working. As the pandemic starts to wind down, there remains a question of paramount importance: will this drastic conversion to remote work last, and if so, to what extent? In this paper, our team examined a variety of variables related to this question, creating several models using elements of applied mathematics, statistics, calculus, economics, and probability.

To begin with, we constructed a model to represent the percentage of jobs by industry that were remote-ready, or able to be operated remotely immediately. Using data from the pandemic and the past, we extrapolated this data to future years through statistical analysis and polynomial regression. Moreover, we developed a COVID factor to account for the pandemic's increase in the percentage of remote-ready jobs. We generated a quadratic regression curve to model the effects of the COVID factor early on as the world adjusted to the unique circumstances, and used derivatives to determine when our parabolic-shaped factor began to resemble a line. This regression allowed us to determine the proportion of jobs in each industry that would be remote-ready in a given year.

However, even if change is feasible, there is no guarantee that people will accept it. In order to address this, we developed a model to predict which workers would choose to work remotely and which employers would give their workers this option. Accounting for the personal circumstances of each worker and employer, this second model determines the probability that a worker whose job is remote-ready will actually make the switch. There are many potential factors, but we featured only the most prominent in our model: children, commute time, education, pay, and business size. Example calculations for different employees are also included to demonstrate how our model functions.

Synthesizing these two models, we developed a model to predict the percentage of workers who will work remotely in a given city. We tested our model on five cities: Seattle, Omaha, Scranton, Liverpool, and Barry. For each city, we analyzed the most prominent industries and the demographics of the inhabitants. From our first model, we found the percentage of jobs in each industry that were remote ready, and from our second model, we determined the percentage of workers in each of the remote-ready jobs who would switch based on their demographics. From our models, we made predictions for the percentage of workers who will work remotely in the years 2024 and 2027. We found that in terms of the impact that remote work will have on each city, Barry was predicted to be the most impacted, followed by Omaha, Seattle, Liverpool, and then Scranton.

Contents

2 Part I: Ready or Not

2.1 Problem Restatement

Given the provided data sets, we identified factors that we believed affected the percentage of remote-ready jobs for a given city, such as age demographic, prominent professions, and education level. From there, we took into account the salient effect of the COVID-19 pandemic on the percentages and how it has forced companies to transform rapidly into working digitally. Incorporating these factors, we developed a model to estimate the current percentage of workers whose jobs are remote-ready. Finally, we extrapolated our model to make predictions regarding this statistic for the years 2024 and 2027, accounting for input changes that occur over time.

2.2 Assumptions

1. **Wages and working hours remain constant.** There was no data provided regarding wage or working hour changes, so we assumed for these to remain the same before, during, and theoretically after the pandemic.

2. **The effects of the COVID-19 pandemic are equal across the globe.** This assumption accounts for the variations in the number of cases and deaths during different time frames in regions throughout the world and makes it considerably more practical for us to evaluate its effects. However, this does not account for the differing responses of countries' people and governments; these factors will show themselves in the results.

3. **Without considering COVID-19, the percent of workers with remote-ready jobs is increasing at a constant rate over time.** This assumption accounts for the natural improvement in the flexibility of technology over time to allow more jobs to be performed at home at a sufficiently efficient level. Possible examples include (1) the use of AI to perform the physical aspects of jobs, thereby allowing more workers to work at home and (2) an increase in the quality of video conference software that would give more users across more occupations the feeling of being present in a far away place while residing at home

4. **The effects of COVID-19 on the percentage of workers who have jobs that are remote-ready kick in March of 2020.** This is when school became closed and the reality of COVID-19 really set in for the majority of students, parents, and workers.

5. **The percentage of workers in the "other services" industry that have remote-ready jobs is equal to the percentage of workers in the entire city across all industries.** "Other services" is an extremely broad set of jobs that simply includes jobs that don't fit into other categories. This is the best way to account for diversity of the "other services" industry as it allows us to construct a linear equation for the final calculated percentage.

2.3 Model Development

We split time into two periods: Pre-COVID and Post-COVID. The concepts in this model can be applied to future pandemics that would force people to stay at home for work, but as we are solely concerned with the COVID pandemic for now, we do this division for the purposes of this model.

2.3.1 Pre-COVID

We used the data given for all 5 cities of the number of workers in each industry in addition to the percentage of jobs for each occupation type that can be done at home effectively. The idea of this is to expand upon the percentages given in "Remote Work Data" to find a percentage of remote-ready workers for each industry, weighting that percentage depending on the number of workers in the particular industry, which is given in "City Employment Data." This is necessary as there are more occupations than industries. This weighting model for Pre-COVID also makes logical sense because hypothetically, if a city were to have only construction workers, since it's extremely difficult to perform construction while at home, the city would have a nearly zero percentage of workers with a remote-ready job. In the following table, occupations are assigned to industries. Note also that we used data from the U.S. Bureau of Labor that outlines employment by detailed occupation to determine within each industry, which occupations are the most popular. This calls for more weighted averages within each industry to develop an overall percentage considering all occupations in an industry.

Industry	Occupation	Number of People Employed (thousands)
Mining, Logging, Construction	Construction and extraction	6,971.10
	Farming, fishing and forestry	1,061.80
Manufacturing	Production	8,766.50
Trade, Transportation, Utilities	Building and grounds cleaning and maintenance	5168.8
	Installation, maintenance and repair	5,847.30
	Transportation and material moving	12,673.20
Information	Architecture and engineering	2,603.00
	Computer and mathematical	5,225.00
	Life, physical and social science	1,444.00
Financial Activities	Business and financial operations	9,422.50
	Sales and related	14,393.60
Professional and Business	Legal	1,328.70
	Management	9,782.30
Education and Health	Education, training and library	9134.7
	Healthcare practitioners and technical	9011.4
	Healthcare support	6837.8
Leisure and Hospitality	Arts, design, entertainment, sports and media	2,662.20
	Community and social service	2,804.40
	Food preparation and service related	11,553.30
	Personal care and service	3,870.40
	Protective service	3,417.20
Government	Office and administrative	19,554.70

We let p_p denote the percentage of jobs in a particular industry that are remote-ready and p_i denote the number of workers in that particular industry and we sum them up for all the industries except the "other services" industry. To account for the "other services" industry, we let p_o denote the proportion of workers in the city that work in "other services."

$$percentage = percentage * p_o + \sum p_i p_p$$

$$percentage = \sum p_i p_p / (1 - p_o)$$

We calculated the percentage for each city to obtain 38 percent in Seattle, 37 percent in Omaha, 30 percent in Scranton, 28 percent in Liverpool, and 47 percent in Barry.

2.3.2 The COVID Factor

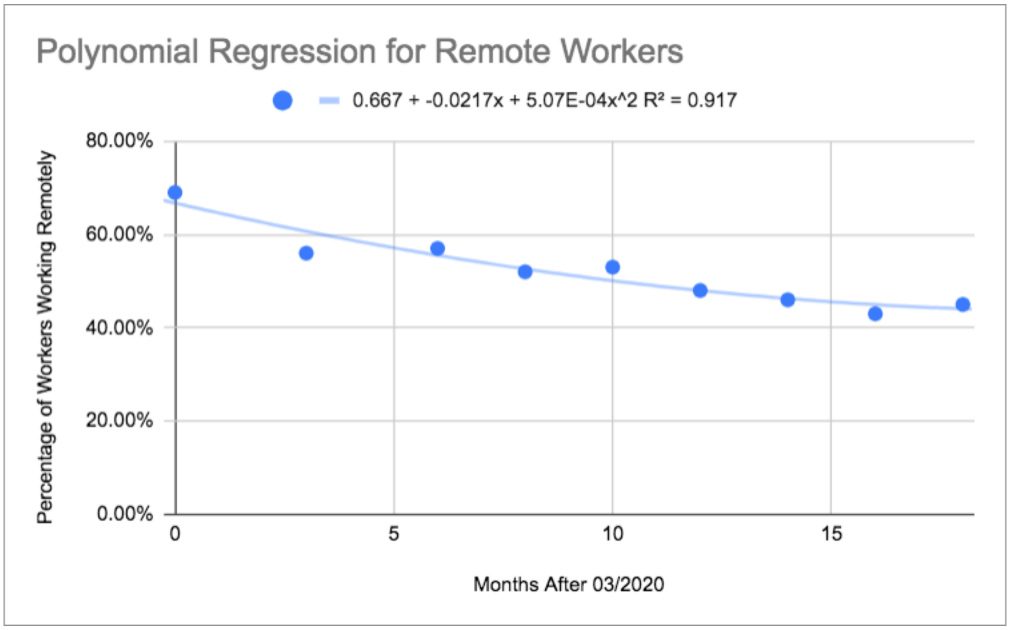
COVID causes a drastic and sudden increase in the percentage of workers who have a remote-ready job. This goes along with our definition of remote-ready jobs as the number of people shifting towards home work suddenly increased due to safety concerns. The most reliable and objective manner to measure our COVID factor f quantity was solely through the education standpoint. We were originally going to use a combination of the education and age factors, but we soon realized that the age factor was almost impossible to quantify reliably because the grouping of age categories in the data sets we tried to combine did not align. For instance, the age demographic categories in Data Set 2 are the median age and the proportions of the population that were under 20 years old as well as between 20 and 29 years old. However, the age categories in data set 3 were uniform age ranges that spanned from a 16-24 to a 65+ category. Thus, the most practical manner in which we could use the age factor was by multiplying the median age by the its demographic's proportion of people that worked a remote-ready job. Solely using the median age statistic for our calculation would have made it inaccurate and exclude large portions of the population. Thus, we decided to exclude age from f. We calculated it by the following equation:

$$f = \sum p_d p_r$$

where p_d is the proportion of the population within the specific age demographic and p_r is the proportion of p_d that work a job that is remote-ready.

Moreover, it is obvious that the effects of the pandemic have been lessened since its introduction in March 2020 [8]. Society begins to adjust to the unique situations and increasingly resemble a normal world as case numbers drop. Thus, the magnitude of the COVID factor lessens as time passes by. To model this, we generated a quadratic regression curve that resembles f in its early stages, and when significant time has passed, we took the derivative of our curve to model the rest of f as linear.

2.4 Results

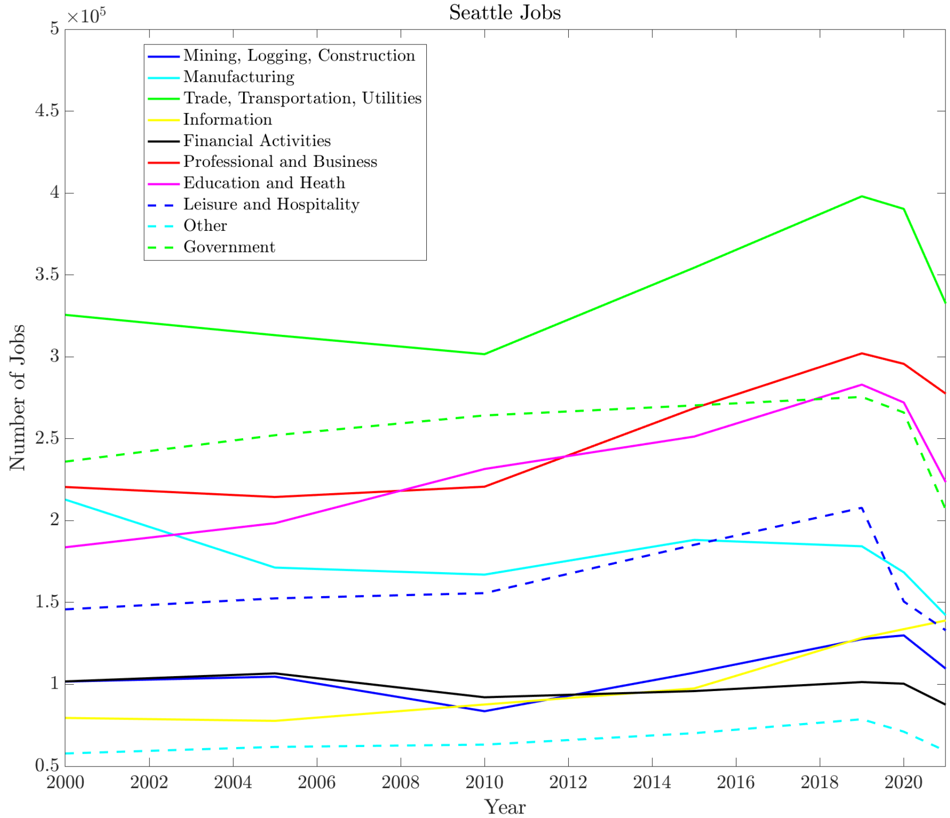


2024 (months after 03/2020)		2027 (months after 03/2020)	
46	54.44%	82	70.71%
47	54.89%	83	71.17%
48	55.35%	84	71.62%
49	55.80%	85	72.07%
50	56.25%	86	72.52%
51	56.70%	87	72.97%
52	57.15%	88	73.43%
53	57.61%	89	73.88%
54	58.06%	90	74.33%
55	58.51%	91	74.78%
56	58.96%	92	75.23%
57	59.41%	93	75.69%
average	56.93%	average	73.20%

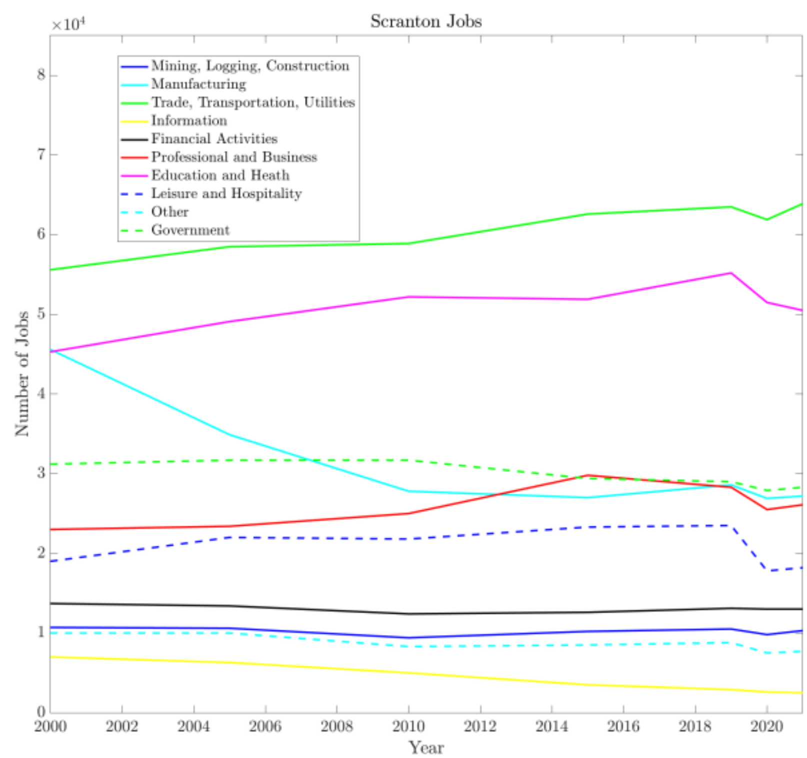
	y = covid factor + pre-covid percentage	0.69 - y	2024 percentage	2027 percentage
seattle	0.6574	0.0326	53.67%	69.94%
omaha	0.6231	0.0669	50.24%	66.51%
scranton	0.5071	0.1829	38.64%	54.91%
liverpool	0.461	0.229	34.03%	50.30%
barry	0.6414	0.0486	52.07%	68.34%

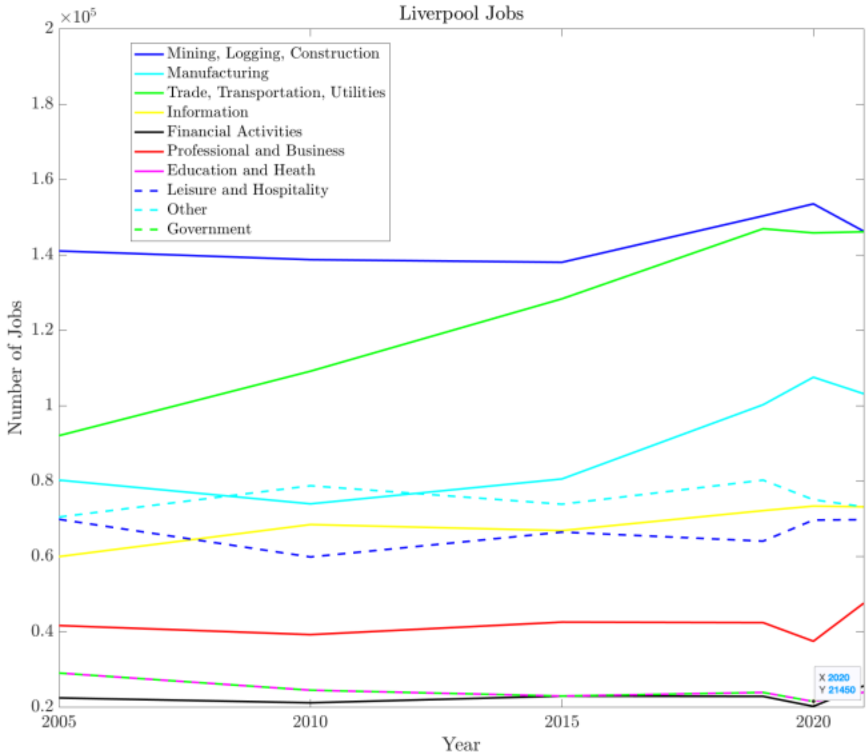
2.5 Strengths and Weaknesses

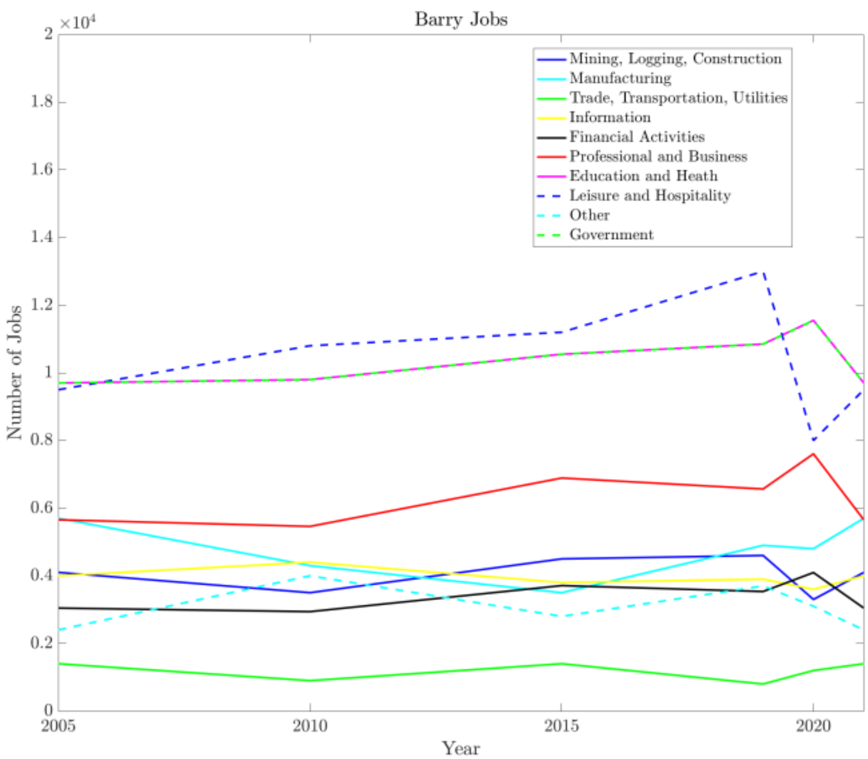
With more data on the trend of cities and their industries, more accurate calculations can be performed. The following graphs are an important visual.

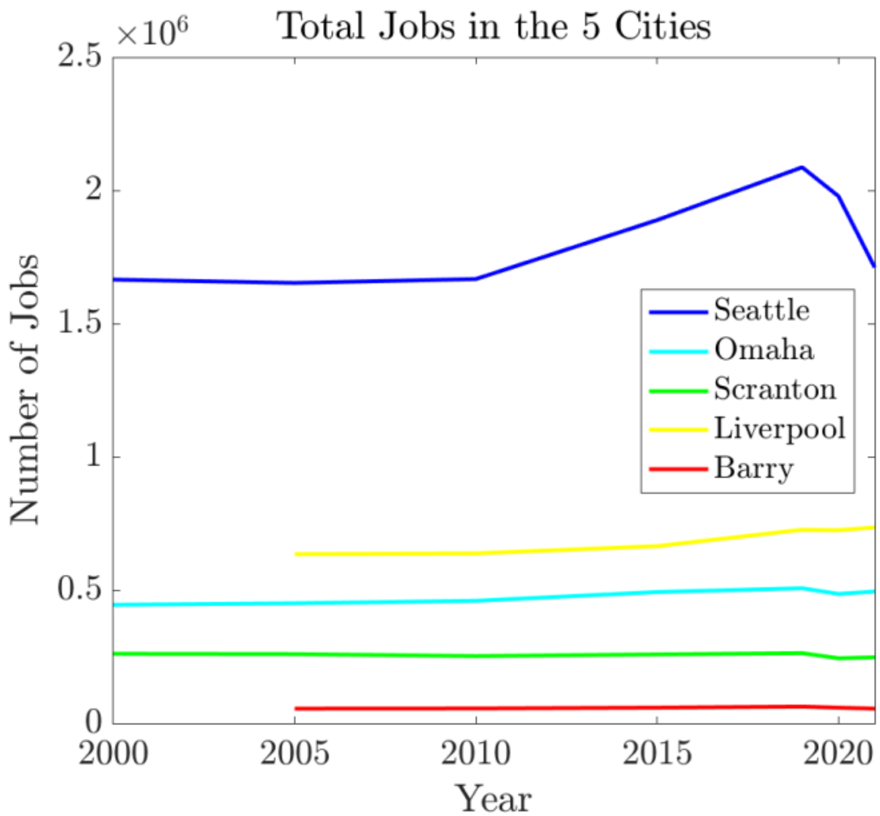



Omaha Jobs
×10^4
Number of Jobs
Year
Mining, Logging, Construction
Manufacturing
Trade, Transportation, Utilities
Information
Financial Activities
Professional and Business
Education and Heath
Leisure and Hospitality
Other
Government


Scranton Jobs
×10^4
Number of Jobs
Year
Mining, Logging, Construction
Manufacturing
Trade, Transportation, Utilities
Information
Financial Activities
Professional and Business
Education and Heath
Leisure and Hospitality
Other
Government


Liverpool Jobs
×10^5
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
Number of Jobs
2005
2010
2015
2020
Year
Mining, Logging, Construction
Manufacturing
Trade, Transportation, Utilities
Information
Financial Activities
Professional and Business
Education and Heath
Leisure and Hospitality
Other
Government
X 2020
Y 21450


Barry Jobs
×10^4
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Number of Jobs
2005
2010
2015
2020
Year
Mining, Logging, Construction
Manufacturing
Trade, Transportation, Utilities
Information
Financial Activities
Professional and Business
Education and Heath
Leisure and Hospitality
Other
Government



3 Part II: Remote Control

3.1 Assumptions

1. **The type of job does not matter as long as it is remote-ready, nor does the age of the employee.** The condition that the job is remote-ready implies that infrastructure is in place to work remotely, meaning there is no job training/additional tech skills required.

2. **The hours and pay remain constant when working from home.** This is a valid assumption because job conditions laid out in contracts usually stipulate wages and hours [4], and the form of working does not affect this [6].

3. **The employer-employee relationship does not affect the probability of being allowed to work from home.** In many jobs, especially large companies, the relation between employer and employee is strictly professional, so there is no emotion interfering with the decision. In the minority of situations where there exists a closer relationship between employer and employee, it can still be ignored for two main reasons. For the purpose of the model as a whole, on average, the effects of emotion should overall be neutral - for every positive relationship, there is a negative relationship. Additionally, it is difficult to quantify an intangible relationship, and in a professional environment, the effects of any positive relationship should be negligible at most.

4. **The gender of the employee does not influence their decision to work remotely, specifically with regards to child-care.** Although females tend to provide more child-care than males, the situation has been moving closer to equality. Also, in most instances, there is paternity leave as well as maternity leave.

5. **The revenue/expenditure of the business is superfluous.** This is because it is accounted for in the size of a business. In particular, businesses with more employees has bigger revenues and bigger expenditures [5].

6. **NQF level is linearly correlated with SAT score.** NQF level and SAT score are both measures of education, and SAT score is strongly correlated with future educational attainment [7]. NQF level is much less quantitatively useful because of unclear mean and standard deviation.

7. **The decision of a worker to work remotely is independent of the employer's decision to allow the worker this option.** It is important to note that although both sides may consider common factors such as young children when making their respective decisions, the final decisions themselves are independent of each other; what the worker personally wishes to do is not influenced by their employer's rules. Since these probabilities are independent, the specific multiplication rule can be utilized to give

$$P_{switch} = P_{choose} * P_{allowed}$$

3.2 Model Development

The two components of our model, as indicated by the above expression, are the independent probabilities that an employee will choose to work from home and that an employee will be allowed to work from home, respectively.

3.2.1 Phase 1: Probability of Choice

In order to make variables consistent, we are defining all of our parameters to be positive and continuous, and their contributions to the overall probability will be reflected in the formula for the probability.

Identification of Variables

Whether or not a worker has children would impact their preference to work at home. In general, employees with children, especially young children, would have a stronger preference to work from home in order to care for them. Since young children require more attention than older children, the impact of young children on the decision to work at home should be greater than older children. Thus, we decided to model the contribution of each child to the child factor, X, as

$$1/(c+1)$$

where c is their age in years. This inverse relationship has a maximum of 1 for a newborn and decreases at a decreasing rate as the age increases; the concave up, decreasing function effectively models the significance of a child's age on the decision to work from home - a newborn has a much larger impact on the decision than a three year old, whereas a 12 year old only has a slightly larger impact than a 15 year old. The total child factor is found by adding all of the individual contributions, or

$$X = \sum \frac{1}{c_i + 1}$$

where c_i is age of the ith child.

The presence of parental leave is a very important consideration for expecting parents. If an employer offers parental leave, there is a smaller incentive to work from home because there will already be time off to care for the child once the baby is born. Furthermore, the factor of parental leave matters more as the due date approaches because it becomes a more imminent issue. We decided to model the pregnancy factor as

$$1/(m + 1)$$

where m is the number of months until the due date. It is similarly modeled to the child factor because the nature and impact of the situation is similar; additionally, this ensures that the graph is continuous at 0 when the baby is born. For the parental leave factor, we modeled it as

$$\sqrt{x/52}$$

where x is the number of weeks of parental leave, an increasing and concave down function. More parental leave is a benefit for expecting parents, which is why the function is increasing, but as the number of weeks of parental leave increases, the marginal benefit of each additional week decreases, which is why the function is concave down; the added benefit of 50 weeks versus 49 is much less than the gain from 1 to 2 weeks of leave.

Similar to parental leave, child care is an important factor for expecting parents and parents of young children, but it does not apply to parents of older children. Specifically, a good indicator of child care is what percentage of preschool is covered. This can be modeled as a negative exponential function with a maximum of 2 and a minimum of 1:

$$2^{1-v/100}$$

where v is the percent of childcare paid for by the employer and/or government. Since we are focusing on paid preschool, we are choosing to only include this term when accounting for children under 7 and for pregnancies. The graph is concave up to show the diminishing marginal returns; as the amount of child care increases, the marginal change is less important.

We decided that for short commute times (under an hour), the model is linear because the marginal impact is consistent for increases in time - for most people, the increased inconvenience of a 30 minute ride compared to a 25 minute ride is the same as that of a 15 minute ride compared to a 10 minute ride. However, above times of one hour, the graph changes to a square root in order to reflect a decreasing marginal inconvenience - for most people, after a certain point, increases to the ride time have a reduced impact because of how long they have already been commuting; they have become desensitized in a way. A model that reflects this is

$$I(t) = \begin{cases} t/60 & t \leq 60 \\ 2\sqrt{t/60} - 1 & t > 60 \end{cases}$$

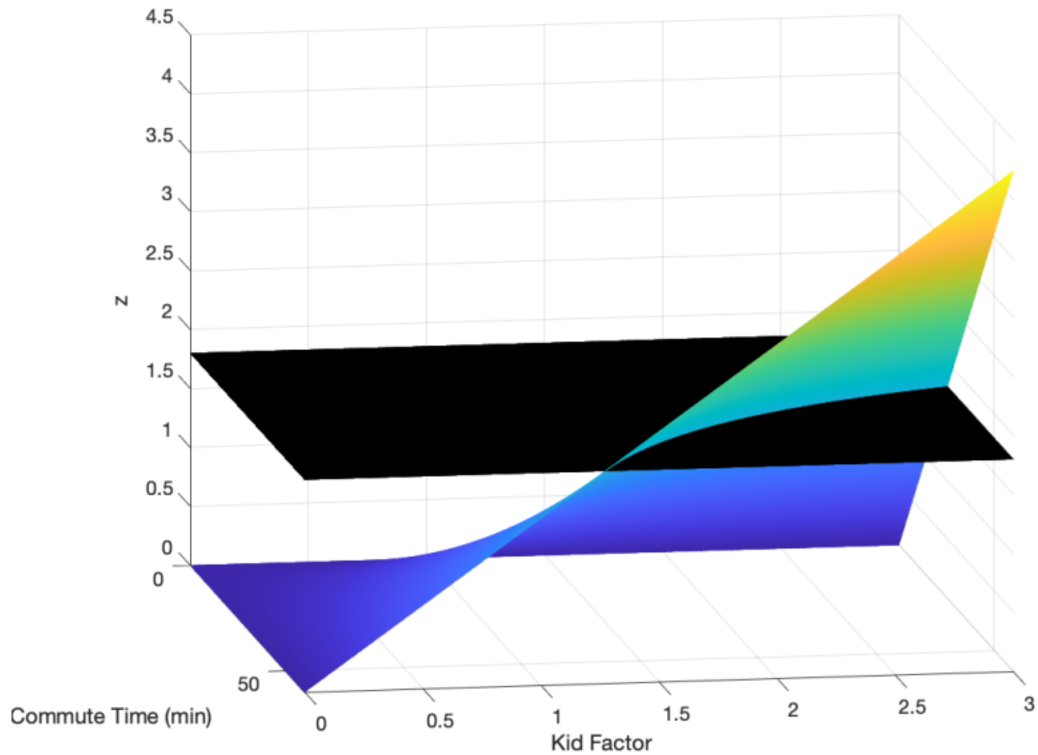
This model is continuous and differentiable at 60, consistent with real world situations. Additionally, the model for commute time also addresses commute cost; in general, commute times are linearly related to commute costs regardless of the method of transportation.

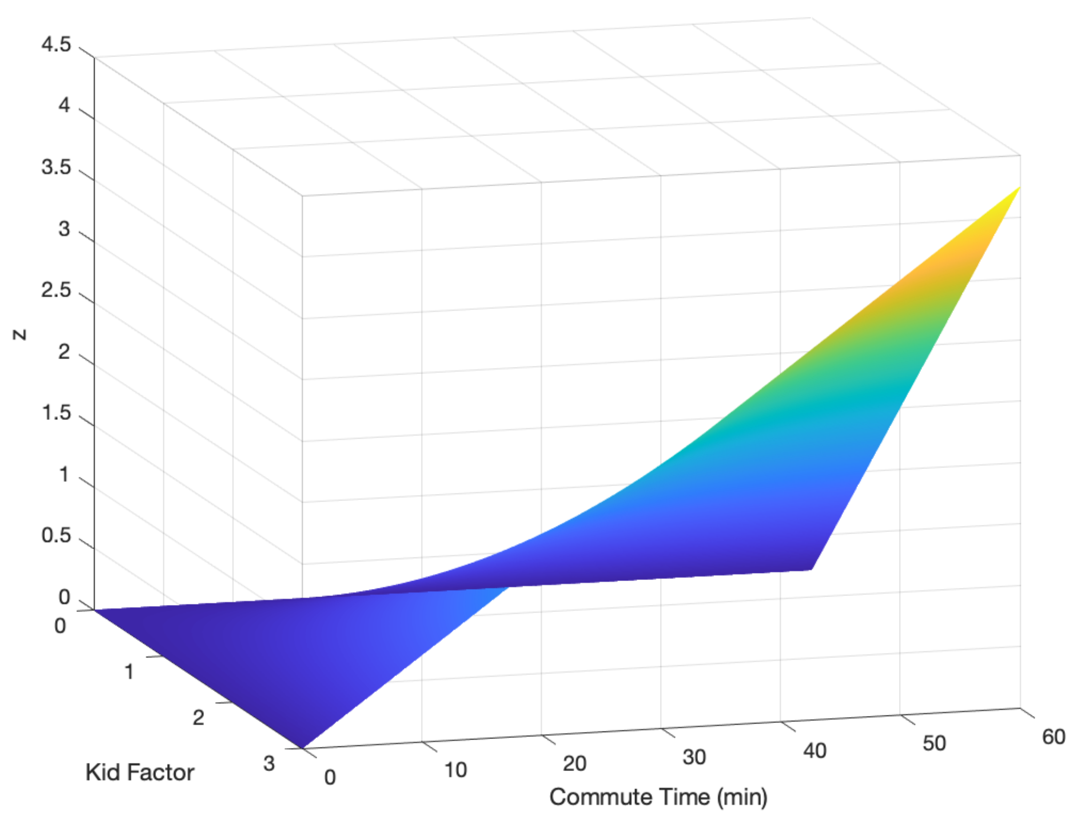
Overall Formula

$$P' = \left(\sum_{c_i<7} \frac{1}{c_i+1}\right) * 2^{1-v/100} + \sum_{c_i \geq 7} \frac{1}{c_i+1} + \left(\sum \frac{1}{m+1}\right) * (1 - \sqrt{x/52}) + I(t)$$

Child-care is most relevant for kids under 7, thus the splitting of kids into those less than 7 and those above. However, for those with kids below 7, the availability of childcare is very important, so the childcare cost factor is multiplied. The index of kids above 7 is included for completeness. When it comes to pregnancy, the situation is intimately tied with childcare. As pregnancy gets closer, the probability of working from home gets larger, but as paternal leave increases, the probability of working from home decreases since it will be possible to spend significant amounts of time with the newborn without staying home. This justifies the negative factor on the parental leave. Finally, a higher commute time incentivizes workers to work from home in order to save time and also money wasted on transportation. Based on a numerical analysis, the maximum reasonable value of this expression is roughly 1.8, as shown by the graphs below. Anything at or above this value should return a probability of 1, and below it, we linearly adjusted the value to satisfy the definition of probability being a number between 0 and 1 inclusive. Specifically, the average commute time is under 30 minutes [3] and the kid factor will be under 2 [2], given the average number of kids and their ages in the US. Plugging in other characteristic factors, the graphs below show the result. Motivated by this, the final probability is

$$P_c = \begin{cases} P'/1.8 & P' \leq 1.8 \\ 1 & P' > 1.8 \end{cases}$$


4.5
4
3.5
3
2.5
z
2
1.5
1
0.5
0
0
50
Commute Time (min)
0
0.5
1
1.5
2
2.5
3
Kid Factor


4.5
4
3.5
3
2.5
z
2
1.5
1
0.5
0
0
1
2
3
Kid Factor
0
10
20
30
40
50
60
Commute Time (min)

3.2.2 Phase 2: Probability of Allowance

In order to make variables consistent, we are defining all of our parameters to be positive and continuous, and their contributions to the overall probability will be reflected in the formula for the probability.

Similar to the employee perspective, the presence of children, especially young children, would be of great importance to an employer. The largest concern associated with remote working is decreased productivity, and having to care for children would likely significantly reduce productivity. As stated previously, young children have a larger impact than older children, meaning their contribution should be weighted more. Thus, the same formula can be used,

$$X = \sum \frac{1}{c_i + 1}$$

Salary is an indicator of how valuable a worker is to the employer. A high salary means the worker is valuable and also irreplaceable; a low wage worker is more likely to be expendable. To an employer, it is advantageous to be more flexible with valuable workers because they are less likely to be easily supplanted. Conversely, denying less skilled workers the option to work from home has less repercussions because if they were to quit, it is likely that a replacement could be found without much hassle. At both ends of the spectrum, small changes in salary are not likely to change the likelihood of being allowed to work from home. On the low end, below a certain threshold, workers are replaceable and thus employers are unlikely to give them alternative options for working; on the high end, above a certain threshold, workers are irreplaceable, and the employer will allow all of these employees greater flexibility such as working from home, so increases in wage will not yield more results once the maximum has been given. The graph that best exemplifies this behavior of a small slope at the end with the largest slope in the middle is the sigmoid graph; for our situation specifically, the hyperbolic tangent graph works best. The mathematical formula is

$$P(s) = \frac{1}{2} \tanh\left(\frac{x - 30}{20}\right) + .5$$

From the data given to us, it appears that in general, educational attainment is normally distributed. Moreover, in general, educational attainment should be normally distributed because standardized tests that filter students to different levels of higher education are intentionally designed to yield normally distributed results [9] [1]. Thus, if the qualifying tests for different levels of educational attainment yield normally distributed results, the levels of attainment themselves should also be normally distributed. The relevant formula pertaining to the probability of allowance is

$$\int_{min}^{e} f \mathrm{d}t$$

where e is the level of educational attainment, min, is the minimum level of educational attainment and f is the (normally distributed) probability distribution. For the justification, consider this: more educated workers are more qualified, meaning in theory, someone with more education should be more qualified than someone with less, meaning that worker has the qualifications for that job and all the ones "below" it in terms of education; i.e. someone with a postgraduate degree should be able to do the tasks required of a "postgraduate job" as well as a "bachelor's degree job", "high school diploma job", and so forth. Since a worker qualified for a certain job should be able to perform all jobs that are "less qualified", this measure of utility of a worker can be represented as the integral of the normal distribution of education attainment, which is an indicator of skills/qualifications. The integral reflects the fact that more educated workers are more qualified and thus able to perform a wider range of jobs, meaning they are more useful to the employer. This result is logically sound and therefore this measure of worker utility is demonstrated appropriately by the integral. As stated in the assumptions, we conflate this with SAT score, which has a minimum of 400, maximum of 1600, mean of 1060 and standard deviation of 217 [1]. Using a min-max normalization on NQF levels, the following table of equivalent scores is obtained.

NQF level	1	2	3	4	5	6	7	8
SAT Score	400	571	743	914	1086	1257	1428	1600

Thus, leveraging the formula for the normal distribution, the distribution of educational attainment is

$$f(t) = \frac{1}{217 * \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-1060}{217}\right)^2}$$

where t ranges from 400 to 1600, the possible range of SAT scores. For the formula, we will integrate f(t) from 400, the minimum SAT score, to s, the equivalent SAT score corresponding to the NQF level obtained through the table above.

A small company would not want their workers to be online because of the costs associated with managing a company from home. The price of online infrastructure and networks are a fixed cost, and with few workers, the cost per worker is too high to justify working from home. On the other hand, a large company with more workers would reduce this cost per worker of working from home, making it more feasible to maintain digital infrastructure. Additionally, with a large workforce, even with a larger percentage of people working online, there would still be enough people working in person to ensure daily upkeep of the physical building space, something that would not be reasonable with a smaller business. The government categorizes business size based on the number of workers, and this scale grows exponentially. Thus, we decided to model this percentage of remote workers as linear with respect to the log of the number of workers:

$$P(w) = \log w$$

where w is the number of workers at the place of employment.

Overall Formula

$$P' = \left(\frac{1}{2}\tanh\left(\frac{x-30}{20}\right) + .5\right) * \left(\int_{400}^{s} \frac{1}{217 * \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t-1060}{217}\right)^2} dt\right) + \left(\frac{\log}{3} w - 2\sum \frac{1}{c_i + 1}\right)$$

The sigmoid function and integral both gives values from 0 to 1 and have synergistic effects (i.e. someone with median education and low salary would have only slightly better chance at being allowed to work remotely than a low education-low salary person, but a median education-median salary person would have much more), so it makes sense to multiply them. Our $\log w$ factor could be significantly greater than 1, so it is used additively. The "child factor" is relatively small for most people so it has a multiplicative factor of 2. The final formula is, correcting for irregularities, is:

$$P_a = \begin{cases} 0 & P' \leq 0 \\ P' & 0 < P' < 1 \\ 1 & P' \geq 1 \end{cases}$$

3.2.3 Phase 3: Combining the Models

As a result of the final assumption, combining the models becomes a simple task of multiplication.

$$P_{overall} = P_c * P_a$$

3.3 Results

To test our model we developed two hypothetical individuals, since no parameters are described in the problem statement. Factors extraneous to our model, such as state, are not described.

3.3.1 Person 1

This person works at McDonald's, has a commute time of 20 minutes, has no kids, is not pregnant, and is offered 12 weeks of parental leave but is not offered any child care benefits. According to our model, their probability of choosing to work from home is 17Additionally, their salary is $\$25,000$, they hold a post-graduate degree (NQF-level 8), and the number of workers at the McDonald's location is 10. This makes the probability that the manager allows this worker to work remotely as 36

Overall, the probability of this worker switching to remote work is 6

This is realistic, since we do not expect this specific worker to be working from home.

3.3.2 Person 2

This person works at Apple, has a commute time of 25 minutes, has two kids of ages 2 and 4, is not pregnant, and is offered 14 weeks of parental leave along with full child care benefits. According to our model, their probability of choosing to work from home is 53Additionally, their salary is $\$100,000$, they hold a post-graduate degree (NQF-level 8), and the number of workers at the Apple location is 200. This makes the probability that the manager allows this worker to work remotely as 100

Overall, the probability of this worker switching to remote work is 53

This is consistent with reality, in that this worker probably has at least the chance to work remotely. In addition, the existence of younger children make this individual significantly more likely to work remotely that the average American.

3.4 Strengths and Weaknesses

3.4.1 Strengths

Our model is particularly strong at accounting for the impact of children on a person's incentive to work remotely. In addition to accounting for the disproportionate impact of younger children compared to older children, we also factored in important job features such as parental leave and child care.

Our model also accounts for education and salary differences between workers and how this affects their likelihood of switching to remote work; more educated and highly paid workers are more likely to choose to be allowed to work remotely. Through our use of the SAT score distribution, we were able to effectively quantify and normalize the distribution of NQT levels which were given but incomplete.

3.4.2 Weaknesses

One weakness that our model has is the inability to distinguish between the race and gender of workers. Issues such as the gender pay gap and racism are difficult to quantify and are thus lacking from our model despite the non-negligible impact that they have.

The probabilities are not exact; we established thresholds using stochastic analysis and rationality to determine when we could prescribe hard limits of 0 or 1 in order to neatly represent all potential outputs as valid probabilities between 0 and 1.

4 Part III: Just a Little Home-Work

4.1 Problem Restatement

4.2 Assumptions

1. **The birth rate is constant.** The years that are to be predicted, 2024 and 2027, are close enough in the future so that any change of the birthrate will be negligible.

2. **All assumptions in Problems 1 and 2 are valid**.

4.3 Model Development

We developed our model by combining the models from parts 1 and 2. From part 1, we analyzed the major industries of each city. This allowed us to predict the percentage of jobs of each industry that would be remote-ready in each of the given years. From model 2, we found the percentage of workers in each industry that would be willing and allowed to work remotely given their demographics and the typical features of the industry. This gives us the total percentage as $\Sigma P_switch * P_industry$

4.4 Results

After analyzing each of the cities, we found that the percentage of workers likely to be working remotely in 2024 was: Seattle: .5367 * .3145 = .168; Omaha: .5024 * 4523 = .2272; Scranton: .3864 * .2375 = .09177; Liverpool: .3403 * .2758 = .0938; Barry: .5207 * .6724 = .3501; and in 2027: Seattle: .6994 * .4102 = .2869; Omaha: .6651 * .5234 = .3481; Scranton: .5491 * .2565 = .1408; Liverpool: .5003 * .2950 = .1476; Barry: .6834 * .6832 = .4669.

4.5 Strengths and Weaknesses

4.5.1 Strengths

Since the cities are differentiated based on both the variables in models 1 and 2, our model incorporates nuance. Specifically, the macro-level changes in model 1 are added on top of the micro-level statistics of model 2. This provides the best picture of how individual workers will react in the greater economy.

4.5.2 Weaknesses

Our solution to this part is very simple, at least on the surface. However, all the variables included in the 1st and 2nd questions are inherent in this model. The numbers we provided are not backed up by work, because we did not have time to include all the calculations. The method taken was to find data about each city pertaining to each of the variables in the second model split by industry, including salary and average commute time.

4.6 Summary

The list of cities in order from most to least impacted by the shift to remote working is Barry, Omaha, Seattle, Liverpool, and Scranton.

References

[1] 2021 Total Group SAT Suite of Assessments Annual Report. URL: reports.collegeboard.org/pdf/2021-total-group-sat-suite-assessments-annual-report.pdf.

[2] Average number of own children... URL: www.statista.com/statistics/718084/average-number-of-own-children-per-family/.

[3] Mike Friedrich. Census Bureau Estimates... URL: www.census.gov/newsroom/press-releases/2021/one-way-travel-time-to-work-rises.html.

[4] Carol Kopp. Terms of Employment. URL: www.investopedia.com/terms/t/terms-of-employment.asp.

[5] Nick Magguilli. What Will Always Be True. URL: www.ofdollarsanddata.com/what-will-always-be-true/.

[6] Jean Murray. What Terms Should Be Included in an Employment Contract? URL: www.thebalancesmb.com/what-terms-should-be-included-in-an-employment-contract-397911.

[7] National SAT Validity Study... URL: satsuite.collegeboard.org/media/pdf/national-sat-validity-study-overview-admissions-enrollment-leaders.pdf.

[8] Remote Work: Fad or Future, MathWorks Math Modeling Challenge 2022. URL: https://m3challenge.siam.org/node/559.

[9] Summary of MCAT Total and Section Scores - AAMC. URL: students-residents.aamc.org/media/8386/download.

5 Appendix

5.1 Part I: Ready or Not

```
% Seattle

df = readtable('seattle.csv','readvariablenames',true, 'variablenamingrule','preserve');

% Data
figure(1)
x = [2000, 2005, 2010, 2015, 2019, 2020, 2021];
y = table2array(df(df.("abc") == 0,2:end));
plot(x, y, 'blue', 'linewidth', 2, 'DisplayName', 'Mining, Logging, Construction');
set(gca, 'fontsize', 18, 'ticklabelinterpreter', 'latex')
legend({}, 'interpreter', 'latex', 'location', 'best')
xlabel('Year', 'interpreter', 'latex')
ylabel('Number of Jobs', 'interpreter', 'latex')
xlim([2000 2021])
ylim([50000 500000])
title('Seattle Jobs', 'interpreter', 'latex')

hold on;
y = table2array(df(df.("abc") == 1,2:end));
plot(x, y, 'cyan','linewidth', 2, 'DisplayName', 'Manufacturing');
y = table2array(df(df.("abc") == 2,2:end));
plot(x, y, 'green','linewidth', 2, 'DisplayName', 'Trade, Transportation, Utilities');
y = table2array(df(df.("abc") == 3,2:end));
plot(x, y, 'yellow','linewidth', 2, 'DisplayName', 'Information');
y = table2array(df(df.("abc") == 4,2:end));
plot(x, y, 'black','linewidth', 2, 'DisplayName', 'Financial Activities');
y = table2array(df(df.("abc") == 5,2:end));
plot(x, y, 'red','linewidth', 2, 'DisplayName', 'Professional and Business');
y = table2array(df(df.("abc") == 6,2:end));
plot(x, y, 'magenta','linewidth', 2, 'DisplayName', 'Education and Heath');
y = table2array(df(df.("abc") == 7,2:end));
plot(x, y, 'blue', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Leisure and Hospitality');
y = table2array(df(df.("abc") == 8,2:end));
plot(x, y, 'cyan', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Other');
y = table2array(df(df.("abc") == 9,2:end));
plot(x, y, 'green', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Government');
hold off;
```

```
% Omaha

df = readtable('omaha.csv','readvariablenames',true, 'variablenamingrule','preserve');

% Data
figure(2)
x = [2000, 2005, 2010, 2015, 2019, 2020, 2021];
y = table2array(df(df.("abc") == 0,2:end));
plot(x, y, 'blue', 'linewidth', 2, 'DisplayName', 'Mining, Logging, Construction');
set(gca, 'fontsize', 18, 'ticklabelinterpreter', 'latex')
legend({}, 'interpreter', 'latex', 'location', 'best')
xlabel('Year', 'interpreter', 'latex')
ylabel('Number of Jobs', 'interpreter', 'latex')
xlim([2000 2021])
ylim([0 150000])
title('Omaha Jobs', 'interpreter', 'latex')

hold on;
y = table2array(df(df.("abc") == 1,2:end));
plot(x, y, 'cyan','linewidth', 2, 'DisplayName', 'Manufacturing');
y = table2array(df(df.("abc") == 2,2:end));
plot(x, y, 'green','linewidth', 2, 'DisplayName', 'Trade, Transportation, Utilities');
y = table2array(df(df.("abc") == 3,2:end));
plot(x, y, 'yellow','linewidth', 2, 'DisplayName', 'Information');
y = table2array(df(df.("abc") == 4,2:end));
plot(x, y, 'black','linewidth', 2, 'DisplayName', 'Financial Activities');
y = table2array(df(df.("abc") == 5,2:end));
plot(x, y, 'red','linewidth', 2, 'DisplayName', 'Professional and Business');
y = table2array(df(df.("abc") == 6,2:end));
plot(x, y, 'magenta','linewidth', 2, 'DisplayName', 'Education and Heath');
y = table2array(df(df.("abc") == 7,2:end));
plot(x, y, 'blue', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Leisure and Hospitality');
y = table2array(df(df.("abc") == 8,2:end));
plot(x, y, 'cyan', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Other');
y = table2array(df(df.("abc") == 9,2:end));
plot(x, y, 'green', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Government');
hold off;
```

```matlab
% Scranton

df = readtable('scranton.csv','readvariablenames',true, 'variablenamingrule','preserve');

% Data
figure(3)
x = [2000, 2005, 2010, 2015, 2019, 2020, 2021];
y = table2array(df(df.("abc") == 0,2:end));
plot(x, y, 'blue', 'linewidth', 2, 'DisplayName', 'Mining, Logging, Construction');
set(gca, 'fontsize', 18, 'ticklabelinterpreter', 'latex')
legend({}, 'interpreter', 'latex', 'location', 'best')
xlabel('Year', 'interpreter', 'latex')
ylabel('Number of Jobs', 'interpreter', 'latex')
xlim([2000 2021])
ylim([0 85000])
title('Scranton Jobs', 'interpreter', 'latex')

hold on;
y = table2array(df(df.("abc") == 1,2:end));
plot(x, y, 'cyan','linewidth', 2, 'DisplayName', 'Manufacturing');
y = table2array(df(df.("abc") == 2,2:end));
plot(x, y, 'green','linewidth', 2, 'DisplayName', 'Trade, Transportation, Utilities');
y = table2array(df(df.("abc") == 3,2:end));
plot(x, y, 'yellow','linewidth', 2, 'DisplayName', 'Information');
y = table2array(df(df.("abc") == 4,2:end));
plot(x, y, 'black','linewidth', 2, 'DisplayName', 'Financial Activities');
y = table2array(df(df.("abc") == 5,2:end));
plot(x, y, 'red','linewidth', 2, 'DisplayName', 'Professional and Business');
y = table2array(df(df.("abc") == 6,2:end));
plot(x, y, 'magenta','linewidth', 2, 'DisplayName', 'Education and Heath');
y = table2array(df(df.("abc") == 7,2:end));
plot(x, y, 'blue', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Leisure and Hospitality');
y = table2array(df(df.("abc") == 8,2:end));
plot(x, y, 'cyan', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Other');
y = table2array(df(df.("abc") == 9,2:end));
plot(x, y, 'green', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Government');
hold off;
```

```matlab
% Liverpool

df = readtable('liverpool.csv','readvariablenames',true, 'variablenamingrule','preserve');

% Data
figure(4)
x = [2005, 2010, 2015, 2019, 2020, 2021];
y = table2array(df(df.("abc") == 0,2:end));
plot(x, y, 'blue', 'linewidth', 2, 'DisplayName', 'Mining, Logging, Construction');
set(gca, 'fontsize', 18, 'ticklabelinterpreter', 'latex')
legend({}, 'interpreter', 'latex', 'location', 'best')
xlabel('Year', 'interpreter', 'latex')
ylabel('Number of Jobs', 'interpreter', 'latex')
xlim([2005 2021])
ylim([20000 200000])
title('Liverpool Jobs', 'interpreter', 'latex')

hold on;
y = table2array(df(df.("abc") == 1,2:end));
plot(x, y, 'cyan','linewidth', 2, 'DisplayName', 'Manufacturing');
y = table2array(df(df.("abc") == 2,2:end));
plot(x, y, 'green','linewidth', 2, 'DisplayName', 'Trade, Transportation, Utilities');
y = table2array(df(df.("abc") == 3,2:end));
plot(x, y, 'yellow','linewidth', 2, 'DisplayName', 'Information');
y = table2array(df(df.("abc") == 4,2:end));
plot(x, y, 'black','linewidth', 2, 'DisplayName', 'Financial Activities');
y = table2array(df(df.("abc") == 5,2:end));
plot(x, y, 'red','linewidth', 2, 'DisplayName', 'Professional and Business');
y = table2array(df(df.("abc") == 6,2:end));
plot(x, y, 'magenta','linewidth', 2, 'DisplayName', 'Education and Heath');
y = table2array(df(df.("abc") == 7,2:end));
plot(x, y, 'blue', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Leisure and Hospitality');
y = table2array(df(df.("abc") == 8,2:end));
plot(x, y, 'cyan', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Other');
y = table2array(df(df.("abc") == 9,2:end));
plot(x, y, 'green', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Government');
hold off;
```

```
% Barry

df = readtable('barry.csv','readvariablenames',true, 'variablenamingrule','preserve');

% Data
figure(5)
x = [2005, 2010, 2015, 2019, 2020, 2021];
y = table2array(df(df.("abc") == 0,2:end));
plot(x, y, 'blue', 'linewidth', 2, 'DisplayName', 'Mining, Logging, Construction');
set(gca, 'fontsize', 18, 'ticklabelinterpreter', 'latex')
legend({}, 'interpreter', 'latex', 'location', 'best')
xlabel('Year', 'interpreter', 'latex')
ylabel('Number of Jobs', 'interpreter', 'latex')
xlim([2005 2021])
ylim([0 20000])
title('Barry Jobs', 'interpreter', 'latex')

hold on;
y = table2array(df(df.("abc") == 1,2:end));
plot(x, y, 'cyan','linewidth', 2, 'DisplayName', 'Manufacturing');
y = table2array(df(df.("abc") == 2,2:end));
plot(x, y, 'green','linewidth', 2, 'DisplayName', 'Trade, Transportation, Utilities');
y = table2array(df(df.("abc") == 3,2:end));
plot(x, y, 'yellow','linewidth', 2, 'DisplayName', 'Information');
y = table2array(df(df.("abc") == 4,2:end));
plot(x, y, 'black','linewidth', 2, 'DisplayName', 'Financial Activities');
y = table2array(df(df.("abc") == 5,2:end));
plot(x, y, 'red','linewidth', 2, 'DisplayName', 'Professional and Business');
y = table2array(df(df.("abc") == 6,2:end));
plot(x, y, 'magenta','linewidth', 2, 'DisplayName', 'Education and Heath');
y = table2array(df(df.("abc") == 7,2:end));
plot(x, y, 'blue', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Leisure and Hospitality');
y = table2array(df(df.("abc") == 8,2:end));
plot(x, y, 'cyan', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Other');
y = table2array(df(df.("abc") == 9,2:end));
plot(x, y, 'green', 'LineStyle', '--', 'linewidth', 2, 'DisplayName', 'Government');
hold off;
```

```
% Total Jobs

df = readtable('all.csv','readvariablenames',true, 'variablenamingrule','preserve');

% Data
figure(6)
x = [2000, 2005, 2010, 2015, 2019, 2020, 2021];
y = table2array(df(df.("city") == 0,2:end));
plot(x, y, 'blue', 'linewidth', 2, 'DisplayName', 'Seattle');
set(gca, 'fontsize', 18, 'ticklabelinterpreter', 'latex')
legend({}, 'interpreter', 'latex', 'location', 'best')
xlabel('Year', 'interpreter', 'latex')
ylabel('Number of Jobs', 'interpreter', 'latex')
xlim([2000 2021])
title('Total Jobs in the 5 Cities', 'interpreter', 'latex')

hold on;
y = table2array(df(df.("city") == 1,2:end));
plot(x, y, 'cyan','linewidth', 2, 'DisplayName', 'Omaha');
y = table2array(df(df.("city") == 2,2:end));
plot(x, y, 'green','linewidth', 2, 'DisplayName', 'Scranton');
y = table2array(df(df.("city") == 3,3:end));
x = [2005, 2010, 2015, 2019, 2020, 2021];
plot(x, y, 'yellow','linewidth', 2, 'DisplayName', 'Liverpool');
y = table2array(df(df.("city") == 4,3:end));
plot(x, y, 'red','linewidth', 2, 'DisplayName', 'Barry');
hold off;
```

5.2 Part II: Remote Control

```
% Mesh Grid

[x, y] = meshgrid(0:.1:60, 0:.005:3)
z = (x / 60) .*  (y * sqrt(2));
mesh(x, y, z);
xlabel ('Commute Time (min)');
ylabel ('Kid Factor');
zlabel ('z');

% z=1.8 plane
hold on;
z = zeros(601, 601);
z = z + 1.8;
surf(x, y, z);
hold off;

view([16.20 -14.60]) % orientation w/o z=1.8 plane

% orientation w/ z=1.8 plane
view([80.83 13.58])
ax = gca;
chart = ax.Children(2);
```

```
int min1 = 1; int max1 = 8; int min2 = 400; int max2 = 1600;

int range1 = max1 - min1;
int range2 = max2 - min2;

double toMultiply = (double) range2 / range1;

for (int i = 0; i <= max1  - min1; ++i)
    System.out.println((int) (min2 + i * toMultiply));
```