Team #15581

Remote Work: Fad or Future

Executive Summary

Mr. President,

In 2019, 5.5% of Americans and 4.7% of U.K. workers were telecommuting. When the Pandemic forced employees to return home, they adapted to working remotely, which still remains today as a popular choice when compared to pre-Pandemic levels. But can this seemingly novel method of working be so widespread and prevalent to maintain its influence in the future?

Our team was tasked with determining whether the seismic shift to remote work will last, and to what extent. Specifically, our focus was on 3 cities across the US (Seattle, Omaha, Scranton) and 2 in the UK (Liverpool and Barry).

The first point of work was realizing that the population could be modeled through a compound interest model, using data of workers in an industry and the significance of an industry in a specific city. Using this adapted model, we were able to calculate the expected percentage of workers currently transitioning to a remote setting, as well their projected levels in the future.

Next, the team needed to figure out a model that would represent if a person who can and is willing to work remotely will take the opportunity and work remotely. This was done through a statistical test that aimed to find what characteristics would be significant to an individual's choice and which statistics would not be. This was put in proportion to a specially derived formula and applied to a few individuals in order to determine if the chance of the individual going remote was above 50% or under, our determined threshold. Above 50% meant likely to go remote and under meant unlikely.

Finally, the team had to combine the 2 models it created previously to estimate the amount of workers that would go virtual in a given city, and how much the city will be affected by such a change. This was done by running a set of 10 years through our first model, separated by 2-to-3-year intervals, and then combining it with the second model into a vector. From this, a magnitude of the vector was determined, and the greater the magnitude of the vector, the greater the impact for individuals going remote would be on the city.

Global Assumptions

1. The work force is comprised of all workers over the age of 16 who are willing and able to work.
    - **Justification:** The work force is defined this way by the US Bureau of Labor Statistics.
2. COVID-19 will continue to affect the economy in the future with the same impact as 2020.
    - **Justification:** COVID-19 permanently altered the economy. For evidence of this change, look no further than the daily labor market.

3. Population/workforce growth rates will stay constant throughout the year (as well as the next 10 years in some cases) and will not be affected by external factors such as legislation or restrictions.
   - **Justification:** Although new legislation is always being passed, the economic landscape has never been altered **severely** by policy change since Franklin Delano Roosevelt's New Deal, or perhaps Lyndon Johnson's Great Society.

Global Definitions

1. "Remote" work: Being able to do all aspects of work completely estranged from a work setting, so a person can do their job anywhere in the world, rather than being limited by the geographical location of their employment.
   - **Justification**: "Shifts in population as people choose where they want to live rather than where their job dictates that they have to live", as taken from the problem statement.
2. "Remote Ready" work: Could work online due to proper personal infrastructure such as computers and telecommunication programs.
   - **Justification**: A person has the resources to work remotely at any moment, however is still not working remotely.

Part 1: Ready or Not

Restatement of the Problem

We were tasked with creating a model estimating the current percentage of remote-ready workers, then predicting the percentages in 2024 and 2027 by applying our model to the following cities:

- Seattle, Washington
- Omaha, Nebraska
- Scranton, Pennsylvania
- Liverpool, England
- Barry, Wales

Local Assumptions

1. There are no external shock factors that will cause significant incompatibility with the current model, such as government policies or regulations.
   - **Justification**: Such shock factors are unpredictable, making it difficult to model its effects.
2. The standardized industry categories listed in Dataset 1 are comparable to the major industry sector subcategories determined by the US Bureau of Labor Statistics.

- **Justification**: Many of these categories, such as financial activities and manufacturing, align between both sources, showing that the distribution of categories is extremely similar.

3. The national average of projected growth rate per industry is applicable to all city examples.

- **Justification**: Industries across the country are connected in the modern economy in such a way that limit intrastate effects.

Symbols Used

$P_{RR:total}$: The percentage of workers whose jobs are currently remote-ready in a city

P_{RR} : The **total** number of remote-ready workers in a particular city and industry

E_i : The number of workers from a particular industry

E_{total} : The **total** number of workers in a particular city

i : Industry category index

r_i : The compound annual rate of change for projected job growth for a certain industry

r_{total} : The compound annual rate of change for projected job growth for a city for all industries

h_i : Estimated percentage of jobs that can be done at home by occupation category

t : Time in years from 2020

Solution and Results

The percentage of workers whose jobs are currently ready-remote in a particular city can be modeled by the simple ratio of:

$$P_{RR:total} = \frac{P_{RR}}{P_{total}}$$

Where P_{RR} is the number of ready-remote jobs for all industries and P_{total} is the total number of jobs in a particular city.

As the workforce of each industry is projected to grow at a constant rate r_i annually across the nation, a compound interest formula can be applied:

$$Compound\ Interest = E_i(1 + \frac{r_i}{n})^{nt}$$

Due to the fact that the model operates on an annual basis:

$$n = 1$$

And as such, the function of a job growth in a certain industry over time of a certain industry can be modeled by the statement:

$$E_i(1 + r_i)^t$$

However, in order to estimate the amount of people who are remote-ready, the percentage of jobs in an industry that can be done remotely must be accounted for, and is represented by h_i.

As a result, the model to represent the total number of remote ready workers in a specific city and industry is:

$$P_{RR} = E_i(1 + r_i)^t * h_i$$

We used a modified compound interest formula as we found data for projected job growth for 2020-2030 from the Bureau of Labor Statistics (BLS). E_i represents the number of workers in 2020 from a particular industry such as Trade, Transportation, and Utilities taken from Dataset 1 [1]. We follow the rest of the formula with $(1 + r_i)^t$ where r_i represents the compound annual rate of change for projected job growth for Trade, Transportation, and Utilities for the years 2020-2030. The data was derived from a Bureau of Labor and Statistics figure which showed some discrepancies regarding industry categories, i, but through the use of weighted averages, we were able to calculate R_i for our standardized industry category list as shown below

i	Standardized Industry Categories(i)	Correlating Industry Categories in Major Industry Sector Table [3]	Compound annual rate of change, 2020–30 (R(i)) (Maximum is 1.0)	Estimated percentage of jobs that can be done at home by occupation category(h(i)) (Maximum is 1.0)
0	Mining, logging, construction	Construction and Mining	0.004876914200	0.00
1	Manufacturing	Goods-producing, excluding agriculture, Manufacturing	0.002243550600	0.01
2	Trade, transportation, and utilities	Utilities, Wholesale trade, Retail Trade, Transportation, and warehousing	0.000345243871	0.03
3	Information	Information	0.010000000000	0.87
4	Financial activities	Financial activities	0.000300000000	0.88
5	Professional and business services	Professional and business services	0.010000000000	0.28
6	Education and health services	Educational services, Health care, and social assistance	0.015702232800	0.98
7	Leisure and hospitality	Leisure and hospitality	0.022000000000	0.26
8	Other services	Community and social service	0.009140863700	0.37
9	Government	Federal government, State, and local government	0.003197973300	0.65

Additionally, this statement was then multiplied by a value h_i which represents the estimated percentage of jobs that can be done at home by occupation category. This was derived from Dataset 3 where we similarly correlated differing industry categories to our standardized industry categories using a one-to-one match. Essentially, if there was an industry which could encompass multiple occupational categories listed in Dataset 3, only one was chosen to eliminate the need to adjust given percentages to account for multiple categories.

To ensure that all industries are accounted for in our final model, we included a summation of all industries, i, which would output the total number of remote-ready jobs for **all** industries. This is given by the formula

$$P_{RR} = \sum_{i=0}^{9} E_i(1+r_i)^t * h_i$$

Finally, we used another modified compound interest rate formula to calculate P_{total}

$$P_{total} = E_{total}(1+r_{total})^t$$

E_{total} represents the total employed workforce in a given city which was determined by adding up the total workforce for every industry under the year 2020, which is the starting time period of our model. r_{total} represents the net grwoth rate of all jobs for a given city, which was obtained through external sources [5][6][7][8][9].

Therefore, our final model can be written as

$$P_{RR:total} = \frac{P_{RR}}{P_{total}} = \frac{\sum_{i=0}^{9} E_i(1+r_i)^t * h_i}{E_{total}(1+r_{total})^t}$$

After extrapolating our function over time for the years 2024 and 2027, we obtained the following table.

City	Percentage of workers whose jobs are currently remote-ready (2022)	Percentage of workers whose jobs are remote-ready (2024)	Percentage of workers whose jobs are remote-ready (2027)
Seattle	0.395	0.382	0.365
Omaha	0.461	0.497	0.555
Scranton	0.205	0.206	0.204
Liverpool	0.293	0.233	0.166
Barry	0.059	0.054	0.048

Short Summary and Interesting Findings

A compound interest formula was implemented to estimate the percentage of a city's workforce that is ready to go remote. From the model, it is very interesting to see which towns are upholding economies that are more reliant on in person attendance, as those economies will go down in remote attendance he more time progresses, and economies that do not rely on in person attendance, will go more remote.

There is a simple explanation for the latter, as going remote is cheaper for businesses who do not have to be concerned with renting/purchasing a space for workers. Additionally, larger cities were more likely to keep going remote, as larger cities tend to be technological centers, and the more technology centered a place, the more of the work can be done remotely.

Model Assessment

Strengths and Weaknesses

In our model, we used data collected in 2020, which was the starting point of the pandemic in which many were left unemployed. Since data provided from other years was not factored in, trends with respect to these previous years were not able to be evaluated.

In the model, the changing total growth of the job market was taken into account, using our modified compound interest formulas in order to estimate the true nature of the remote readiness. This is important due to the fact that as a labor force increases increases, the amount of people within that labor force who will work remotely increases as well. However, if we keep the remote projections in proportion to the expected workforce, it allows to draw conclusions about the true nature of working remotely in a city.

Additionally, the model is important in determining the general trends regarding a city's economic makeup, as a metropolitan city is more opportunistic towards technological jobs. These can be done remotely compared to jobs in a small town, which will more often have jobs that are difficult to perform remotely.

When there were discrepancies with industry categories across 3 different datasets, we were able to make a standardized list and use weighted averages to create a reliable and accurate representation of a job sector's growth. This was important, as it more accurately represents the economic makeup of a city in the model.

Part 2: Remote Control

Restatement of the Problem

We were tasked with creating a probabilistic model to determine whether a person will work remotely given that they are remote-ready.

Local Assumptions

1. In order for the model to function, age, sex, race, marriage, childcare, and education are the principal determinants of working remotely.
    - **Justification**: These were values considered by the US Bureau of Labor Statistics [2].
2. The probability of people working remotely given that they are remote-ready in June 2020 is similar to that of prior and future months and years.
    - **Justification**: People that are remote-ready will commit to the workplace change overtime due to the plethora of benefits, including more sleep time and money conserved from commuting. Thus, if a

person is remote-ready, an assumption can be made that their likelihood of becoming remote is not time-dependent.

3. The data from the US is representative of the combined population of the US and the UK.
    - **Justification:** The US and the UK are both capitalist, modernized countries with democratic governments, and similar philosophies when implementing economic policy. Due to the industries overlapping significantly, and those industries being regulated in much the same way, data from the US can be extrapolated to the UK as well.

Symbols Used

$\overrightarrow{d_O}$: Vector with observed distributions

$\overrightarrow{d_E}$: Vector with expected distributions

$\overrightarrow{X^2}$: Vector with Chi-Square test statistics

$\overrightarrow{df}$: Vector with degrees of freedom

α : Our alpha value

$\overrightarrow{p}$: Vector with p-values

$P(R)$: Probability of working remotely

R_A : The percent of people who teleworked depending on age

R_S : The percent of people who teleworked depending on sex

R_R : The percent of people who teleworked depending on race

R_M : The percent of people who teleworked depending on marriage

R_C : The percent of people who teleworked depending on childcare

R_E : The percent of people who teleworked depending on education

$P(RR)$: Probability of being remote-ready

E_i : The population of workers per industry in 2020 for the entire US

h_i : Estimated percentage of jobs that can be done at home by occupation category

E_{total} is the total US workforce

$P(R \cap RR)$: Probability of working remotely and being remote ready

$P(R|RR)$: Probability of working remotely given that a person is remote ready

Note: R_A, R_S, R_R, R_M , R_C, and R_E are all situational values from $\overrightarrow{d_O}$ for a specific person

Solution and Results

Justification of Indicators

To see whether a variety of variables were important indicators of remote working, we used Chi-Square Goodness of Fit tests. We generalized these tests to each element in our vector for each variable; our observed distribution was the percentages of remote workers by category in June 2020 [3]. As such, $\overrightarrow{d_O}$ was

$$\begin{gathered}\langle\{15, 35.6, 35.6, 33, 30.5, 27\},\\ \{27.2, 36\},\\ \{30.8, 25.7, 48.5, 21.1\},\\ \{35, 27.9, 26.3\},\\ \{34.7, 29.6\},\\ \{33.4, 4.8, 12.6, 22.3, 54.1\}\rangle\end{gathered}$$

For each variable, our expected distribution was a uniform distribution encapsulating the overall percentage of remote workers in June 2020 [3]. For the first 5 vector elements, we used the percentage of remote workers aged 16 and over. For the last vector element, we used the percentage of remote workers aged 25 and over. As such, $\overrightarrow{d_E}$ was

$$\begin{gathered}\langle\{31.3, 31.3, 31.3, 31.3, 31.3, 31.3\},\\ \{31.3, 31.3\},\\ \{31.3, 31.3, 31.3, 31.3\},\\ \{31.3, 31.3, 31.3\},\\ \{31.3, 31.3\},\\ \{33.4, 33.4, 33.4, 33.4, 33.4\}\rangle\end{gathered}$$

Our null hypothesis (H_0) for each vector component was that $d_{O_i} = d_{E_i}$ and our alternate hypothesis was that $d_{O_i} \neq d_{E_i}$. The following conditions were satisfied:

1. The variables were strictly categorical.
2. The expected value for each comparison was at least 5.
3. The data was approximately random. (Our values were actually obtained from a June 2020 US census, but it could be viewed as a random sample with regards to the combined US and UK populations.)

We used the following formula to find our X^2 values:

$$X_i^2 = \sum_{j=0}^{n} \frac{(d_{O_{ij}} - d_{E_{ij}})^2}{d_{E_{ij}}}$$

$\overrightarrow{X^2}$ was $\langle 10.4, 1.24, 13.8, 1.61, .462, 54.0\rangle$.

Then, we calculated the degrees of freedom using the following formula with respect to the cardinality, or number of elements, of the expected distribution vector:

$$df_i = \overline{\overline{d_{E_i}}}$$

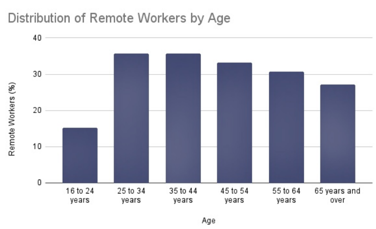
$\overrightarrow{df}$ was $\langle 5, 1, 3, 2, 1, 4 \rangle$.

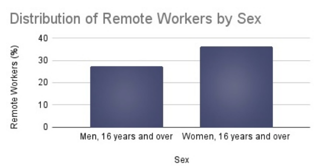
Finally, we found our p-values by applying the following formula, with an $\alpha = .1$:

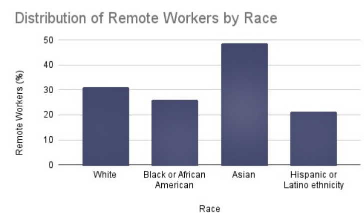
$$p_i = P(H_0)$$

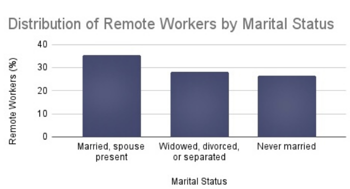
As a result, $\overrightarrow{p}$ was $\langle .065, .26, .0032, .45, .50, 0 \rangle$.

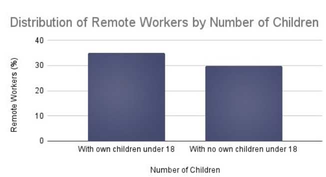
We obtained 3 significant p-values (age, race, and education). Although the other p-values (sex, marital status, and number of children) were not initially significant, they all had an abnormally small number of categories. As such, we used histograms to indicate that they had weaker levels of significance. Below, find the histograms for all our data:

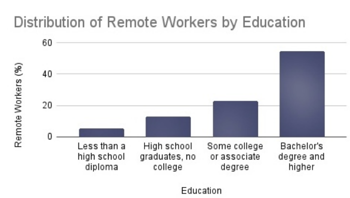












Calculating Probability of Working Remotely

To find the probability of working remotely, we developed the following equation:

$$P(R) = \frac{2 * (R_A + R_R + R_E) + 1 * (R_S + R_M + R_C)}{(2 * 3) + (1 * 3)}$$

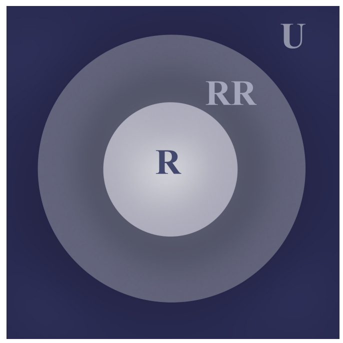
Our equation is a weighted average of probabilities based on 6 variables (age, sex, race, marital status, number of children, and education). The stronger statistically significant factors were given a weight 2 times more than the weaker statistically significant factors.

Calculating Probability of Working Remotely Given Ready-Remote

When finding the probability of a person working remotely given ready-remote, the reasoning can be developed based on cities (Question 1), just applied to the national level. Due to this, the following equation was developed:

$$P(RR) = \frac{\sum_{i=0}^{9} E_i * h_i}{E_{total}}$$

We know $P(R \cap RR) = P(R)$ because R is a subset of RR. Visually this looks like this:



Then, we can calculate conditional probability using the formula:

$$P(R|RR) = \frac{P(R \cap RR)}{P(RR)}$$

Combining each of the previous steps and applying the transitive postulate, we obtain:

$$P(R|RR) = \frac{P(R)}{P(RR)} = \frac{\frac{2*(R_A+R_R+R_E)+(R_S+R_M+R_C)}{9}}{\frac{\sum_{i=0}^{9} E_i*h_i}{E_{total}}}$$

$$P(R|RR) = \frac{[2*(R_A + R_R + R_E) + (R_S + R_M + R_C)] * E_{total}}{9 * \sum_{i=0}^{9} E_i * h_i}$$

Sample Calculations

To demonstrate the efficacy of this model, we applied it to the following cases with results as shown in the table below.

Case	Value	Boolean (to go remote)
26 year old, male, Asian, never married, no children, Bachelor’s degree	0.631	Yes
62 year old, female, Black, married, 3 children, masters degree	0.574	Yes

45 year old, female, White, married, 2 children, high school graduate	0.454	No

To find the $P(R|RR)$ of example 1, we plugged in the percent of people who worked remote for each factor in the $P(R)$. Additionally, we used national data for each industry [4]. Therefore,

$$P(R|RR) = \frac{P(R)}{P(RR)} = \frac{\frac{2*(R_A+R_R+R_E)+(R_S+R_M+R_C)}{9}}{\frac{\sum_{i=0}^{9} E_i*h_i}{E_{total}}} = \ .631$$

We assume that a value greater than 0.5 will imply the individual will choose to take a remote job and a value less than 0.5 will imply the individual will not choose to take the remote job. In this case, the 26 year old male took the remote job.

Model Assessment

The model is subjective in its assigning of weight in statistically significant and insignificant variables (a subjective 2 to 1 weight respectively was chosen), skewing the results. If research is done and values are obtained, the model will become even more accurate. Another weakness is our 0.5 threshold, which is chosen due to the Boolean effect that it provides (either Yes or No). There is no mathematical derivation of this value.

One of the strengths of our model is our use of 6 unique variables, namely, age, race, education, sex, marital status, and number of children. This usage of multiple variables was strengthened through our weighted average said variables through the statistical analysis of all factors using Chi-Square Goodness of Fit tests. By defining multiple factors, we were able to assign multiple characteristics to an individual in a test case, which allowed for redundancy within the model in case some of the data was skewed.

Part 3: Just a Little Home-work

Restatement of the Problem

We were asked to synthesize our first two solutions to create a comprehensive model to estimate the percentage of remote workers in a particular city, then rank the cities from Question 1 with respect to the magnitude of impact of remote work.

Local Assumptions

1. The only 2 significant factors for this model are age and education.
   - **Justification:** Other data was not available to consider in this case, if it was, the model would be immensely powerful and accurate.
2. The averages in this model were based on approximately normal distributions. That being said, the distributions were not exactly normal.

3. The magnitude of impact that remote working will have is defined as the city with the greatest percentage of remote working over time. We could quantify this value using a vector magnitude (see below).
    - **Justification:** We interpreted the magnitude of remote working to be the amount of people who are choosing this lifestyle.

Symbols Used

R_A : The percent of people who teleworked depending on age

R_E : The percent of people who teleworked depending on education

E_i : The number of workers for a particular industry

E_{total} : The **total** number of workers in a particular city

i : Industry category index

h_i : Estimated percentage of jobs that can be done at home by occupation category

t : Time in years from 2020

r_i : The compound annual rate of change for projected job growth for a certain industry

d_{O_e} : Observed distribution based on category of industry

d_{E_e} : Expected distribution based on category of industry

X_e^2 : Chi-Square test statistic for industries

df_e : Degrees of freedom for industries

α : Our alpha value for industries

p_e : P-value for industries

$P(RR)_e$: Probability of being remote-ready contingent on employment

$P(R|RR)$: Probability of being remote given that a person is remote-ready

$P(R)_e$: Probability of being remote contingent on employment

$P(R)_p$: Probability of being remote contingent on the person

- Justification: The "person" is the sum of their attributes. In Question 2, this included 6 factors. In Question 3, this will include 2 factors (age, education) that are the easiest to quantify in a limited timespan.

Solution and Results

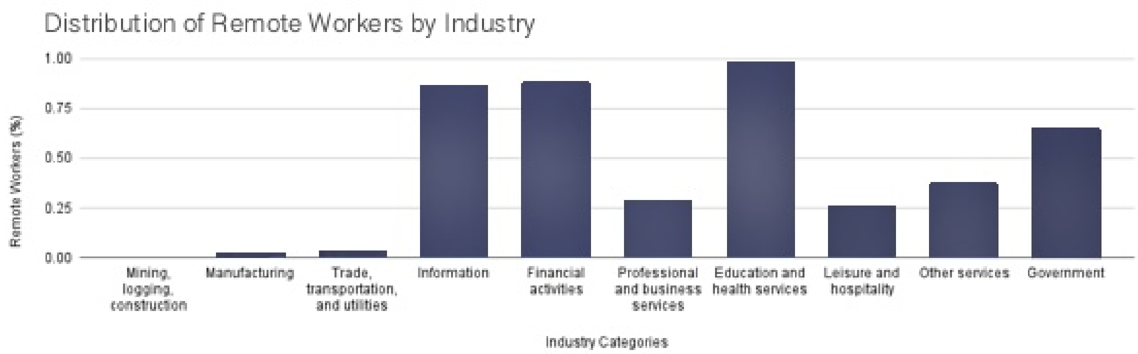
Justification of Indicators

To see whether industry was an important indicator for remote working, we ran another Chi-Square Goodness of Fit test. d_{O_e} was $\{0, 1, 3, 87, 88, 28, 98, 26, 37, 65\}$ and d_{E_e} was $\{100, 100, 100, 100, 100, 100, 100, 100, 100, 100\}$. We assumed 100% for all values to indicate that remote working was not of importance. As such, our null hypothesis (H_0) for each vector component was that $d_{O_i} = d_{E_i}$ and our alternate hypothesis was that $d_{O_i} \neq d_{E_i}$. Again, the following conditions were satisfied:

1. The variables were strictly categorical.
2. The expected value for each comparison was at least 5 (almost true).
3. The data was approximately random.

Then, $X_e^2 = 453.81$. Again, assuming $\alpha = .1$ and finding $df_e = 9, p_e = 0$. Since our p-value was less than our alpha-value, we could conclude that industry was statistically significant.

We generated the following histogram:



We did not consider location to be an indicator, since location will only influence the initial population of remote workers. Additionally, it will factor into the person's assessment and be directly linked. Any influence prohibits testing using a Chi-Square distribution.

Calculating Probability of Working Remotely in a Given City

$$P(R)_{e,p}(t) = \frac{2 * P(RR)_e \; * \; P(R|RR) \; + \; 9 * P(R)_p}{11}$$

Again, we computed a weighted average. By multiplying $P(RR)_e$ by $P(R|RR)$, we obtained $P(R)_c$. We multiplied $P(R)_c$ by 2 since the industry was statistically significant. We multiplied $P(R)_p$ by 9 since we had 6 variables. 3 of these variables were statistically significant and the other 3 had a weaker significance. Then, we could divide by 2+9=11.

Our final aggregate model is:

$$P(R)_{e,p}(t) = \frac{2 * \frac{\sum_{i=0}^{9} E_i (1+r_i)^t * h_i}{E_{total} (1+r_{total})^t} * \frac{[2*(R_A+R_R+R_E)+(R_S+R_M+R_C)]*E_{total}}{9*\sum_{i=0}^{9} E_i * h_i} + \frac{2*(R_A+R_R+R_E)+(R_S+\mathrm{R_M} + \mathrm{R_C})}{1}}{11}$$

In our simplified Question 3, each person was only assessed based on their education and age. So, our truncated model was:

$$P(R)'_{e,p}(t) = \frac{P(RR)_e \ * \ P(R|RR) \ + \ 2 * P(R)'_p}{3}$$

Our final and simplified aggregate model is:

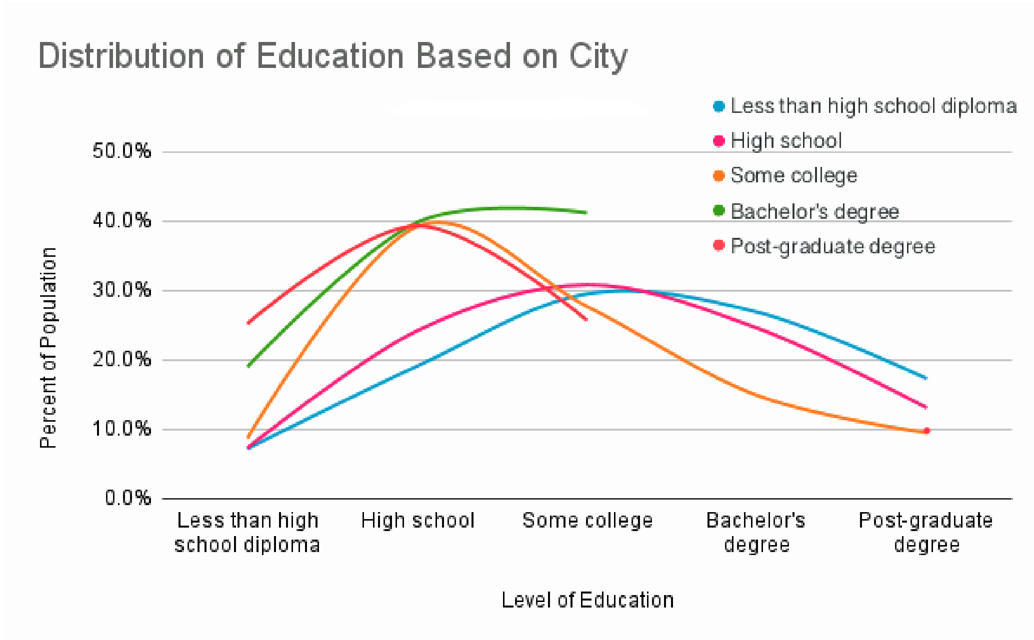
$$P(R)'_{e,p}(t) = \frac{\frac{\sum_{i=0}^{9} E_i(1+r_i)^t * h_i}{E_{total}(1+r_{total})^t} \ * \ \frac{(R_A+R_E)*E_{total}}{2*\sum_{i=0}^{9} E_i * h_i} + \ \frac{R_A+R_E}{2}}{3}$$

Ranking of the Cities

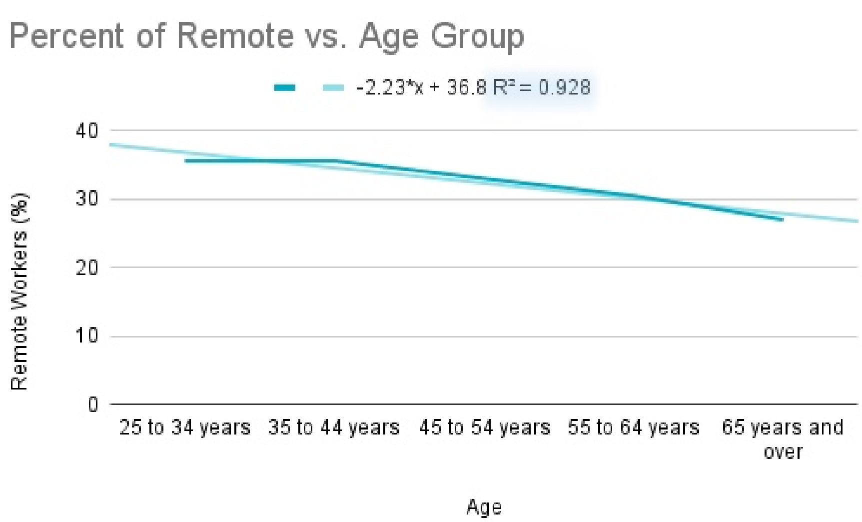
Using our simplified model, we ranked the cities by:

1. Calculating $P(R)'_{e,p}(2)$, $P(R)'_{e,p}(4)$, $P(R)'_{e,p}(7)$, $P(R)'_{e,p}(10)$ for each city.
2. Organizing these values into 5 vectors: $\overrightarrow{P(R)'_{seattle}}$, $\overrightarrow{P(R)'_{omaha}}$, $\overrightarrow{P(R)'_{scranton}}$, $\overrightarrow{P(R)'_{liverpool}}$, $\overrightarrow{P(R)'_{barry}}$ each containing 4 components.
3. Finding the magnitude of each vector and ranking the magnitudes ordinally.

City	Percentage of workers who are projected to be remote (2022)	Percentage of workers who are projected to be remote (2024)	Percentage of workers who are projected to be remote (2027)	Percentage of workers who are projected to be remote (2030)
Seattle	0.395	0.382	0.365	0.348
Omaha	0.461	0.497	0.555	0.620
Scranton	0.205	0.206	0.204	0.204
Liverpool	0.293	0.233	0.166	0.119
Barry	0.059	0.054	0.048	0.043



Because we had a normal distribution for education in each city, we could use the median or mean values to represent an "average". Then, we used this average education level and extrapolated each value to data table [3] which gave us R'_e values for every city as shown below. For our R'_a value, we attempted to use the same technique as the R'_e, but such a model was deemed too insensitive. Therefore, we created a linear regression with a good fit, which correlated the age of remote workers related to the percent unemployment. With this method, we could bypass using the median and obtain more sensitive values:



City	Age	Education
Seattle	0.318717	0.223
Omaha	0.320947	0.223
Scranton	0.306898	0.126
Liverpool	0.31671	0.223
Barry	0.31671	0.126

City	P(R||RR)
Seattle	.429
Omaha	.430
Scranton	.342
Liverpool	.427
Barry	.350

Using the above tables and final model, we were able to find the the percentage of workers who will work remotely given years as shown below.

City	2024	2027
Seattle	0.145	0.142
Omaha	0.162	0.170
Scranton	0.096	0.095
Liverpool	0.123	0.114
Barry	0.080	0.079

So,

$$\overrightarrow{P(R)'_{seattle}} = \langle 0.147, 0.145, 0.142, 0.140 \rangle,$$
$$\overrightarrow{P(R)'_{omaha}} = \langle 0.157, 0.162, 0.170, 0.180 \rangle,$$
$$\overrightarrow{P(R)'_{scranton}} = \langle 0.096, 0.096, 0.095, 0.095 \rangle,$$
$$\overrightarrow{P(R)'_{liverpool}} = \langle 0.132, 0.123, 0.114, 0.107 \rangle,$$
$$\overrightarrow{P(R)'_{barry}} = \langle 0.081, 0.080, 0.079, 0.079 \rangle.$$

The vector magnitude is used in the field of linear algebra to summarize the numeric value for a vector. A vector can be thought of as similar to a 1D matrix or a collection of related values. Taking inspiration from this field of mathematics, we summarized the impact of remote working on a city. The formula for the magnitude of a vector is:

$$\overrightarrow{P(R)'_{city}} = \sqrt{P(R)'^{\,2}_1 \ + \ P(R)'^{\,2}_2 \ + \ ... \ P(R)'^{\,2}_n}$$

The magnitude of each of the 5 vectors is:

$|P(R)'_{seattle}| = .29$ $|P(R)'_{omaha}| = .33$ $|P(R)'_{scranton}| = .19$

$|P(R)'_{liverpool}| = .24$ $|P(R)'_{barry}| = .16$

Ultimately, our ranking of the cities for the magnitude of impact that remote working will have is:

All in all, Omaha will be the most impacted and Barry will be the least impacted.

Model Assessment

In our model, we did not account for races and education can be quantified differently in the age of online education. For example, many individuals can choose to browse free YouTube videos to augment their education and even launch their own businesses before college. This was one notable weakness, exacerbated by the fact that we had to use a truncated model in this final model. In the future, we would like to factor in more data in our model.

Overall, our model is a great synthesis of our Question 1 and 2 solutions. Our model is statistically sound and based on a multivariable equation. It successfully estimates the percentage of workers that will work remotely in any city by taking into account their personal identity and occupation. It enabled us to make reasonable predictions for remote working in 2024 and 2027, and rank the cities in terms of magnitude of the impact remote working would have.

References

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

[2] U.S. Bureau of Labor Statistics. (2020, June). *Ability to work from home:*

Evidence from two surveys and implications for the labor market in the COVID-19 pandemic : Monthly Labor Review. U.S. Bureau of Labor Statistics. Retrieved from https://www.bls.gov/opub/mlr/2020/article/ability-to-work-from-home.htm

[3] U.S. Bureau of Labor Statistics. (n.d.). *Data tables (XLSX)*. U.S. Bureau of Labor Statistics. Retrieved February 28, 2022, from https://www.bls.gov/cps/effects-of-the-coronavirus-covid-19-pandemic.htm#data

[4] U.S. Bureau of Labor Statistics. (2021, September 8). *Employment by Major Industry Sector*. U.S. Bureau of Labor Statistics. Retrieved from https://www.bls.gov/emp/tables/employment-by-major-industry-sector.htm#top

[5] *2020 workforce equity update report - seattle*. (n.d.). Retrieved from https://www.seattle.gov/Documents/Departments/HumanResources/Workforce%20Equity/2020%20WFE%2

[6] U.S. Bureau of Labor Statistics. (n.d.). *Omaha*. U.S. Bureau of Labor Statistics. Retrieved from https://www.bls.gov/regions/midwest/ne_omaha_msa.htm

[7] Economy in Scranton, Pennsylvania. (n.d.). Retrieved from https://www.bestplaces.net/economy/city/pennsylvania/scranton#:~:text=Scranton has seen the job, the US average of 33.5%25.

[8] *Liverpool LGA*. Employment status | Liverpool City Council | Community profile. (n.d.). Retrieved from https://profile.id.com.au/liverpool/employment-status#:~:text=Between%202011%20and%202016%2C%20the,increase%20of%2010%2C481%20or%2013.

[9] *Labour Market Statistics (annual population survey): April 2020 to March 2021*. GOV.WALES. (n.d.). Retrieved from https://gov.wales/labour-market-statistics-annual-population-survey-april-2020-march-2021-html