Dawson Tam 2023-07-06 **Abstract** In an effort to undertsand the impact of Covid-19 on the world, I will analyze the multiple facets of covid related deaths and compare them across variables I deem relevant. Additionally, I want to measure whether or not states with higher vaccine rates see a lower death rate to prove the effectiviness of vaccines. library(tidyverse) ------ tidyverse 2.0.0 --## — Attaching core tidyverse packages —— ## **✓** dplyr 1.1.2 **✓** readr 2.1.4 ## **✓** forcats 1.0.0 **✓** stringr 1.5.0 ## **✓** ggplot2 3.4.2 **✓** tibble 3.2.1 ## **✓** lubridate 1.9.2 **✓** tidyr 1.3.0 ## **✓** purrr 1.0.1 ## — Conflicts —— ——— tidyverse_conflicts() — ## * dplyr::filter() masks stats::filter() ## * dplyr::lag() masks stats::lag() ## i Use the conflicted package (http://conflicted.r-lib.org/) to force all conflicts to become errors library(dplyr) library(skimr) library(Hmisc) ## Attaching package: 'Hmisc' ## The following objects are masked from 'package:dplyr': src, summarize ## ## The following objects are masked from 'package:base': format.pval, units library(ggplot2) summarizing the global data to see which parts of the data sheet need cleaning and processing Covid_19_rates <- read.csv("/Users/dawsontam/Desktop/COVID19_line_list_data.csv")</pre> describe(Covid_19_rates) ## Covid_19_rates ## ## 27 Variables 1085 Observations ## ------## id n missing distinct Info Mean Gmd .05 .10 1085 0 1085 1 543 362 55.2 109.4 ## .25 .50 .75 .90 .95 ## 272.0 543.0 814.0 976.6 1030.8 ## lowest: 1 2 3 4 5, highest: 1081 1082 1083 1084 1085 ## -----## case_in_country n missing distinct Info Mean . 05 .10 ## 4.00 197 197 48.84 2.00 ## .50 . 75 .90 . 25 . 95 ## 11.00 28.00 67.25 110.30 153.65 ## ## lowest: 1 2 3 4 5, highest: 365 443 875 925 1443 ## -----## reporting.date n missing distinct ## 1084 1 ## ## lowest : 02/01/20 02/02/20 02/03/20 02/04/20 02/05/20 ## highest: 2/24/2020 2/25/2020 2/26/2020 2/27/2020 2/28/2020 ## -----## summary n missing distinct ## 1080 5 967 ## lowest : confirmed COVID-19 pneumonia patient No.11 in Tianjin: female, 55, symptom onset on 01/23/2020, hospi talized on 01/23/2020, confirmed on 01/26/2020 confirmed COVID-19 pneumonia patient No.12 in Tianjin: female, 79, symptom onset on 01/24/2020, hospitalized on 0 1/24/2020, confirmed on 01/26/2020 confirmed COVID-19 pneumonia patient No.13 in Tianjin: female, 19, symptom onset on 01/19/2020, hospitalized on 0 1/20/2020, confirmed on 01/26/2020 confirmed COVID-19 pneumonia patient No.14 in Tianjin: male, 71, Wuhan resident, visited Malaysia from 01/19/2020 to 01/25/2020, arrived in Tianjin on 01/25/2020, symptom onset on 01/25/2020, hospitalized on 01/25/2020, confirm ed on 01/26/2020 confirmed imported COVID-19 pneumonia patient in Gansu: female, 20, lives in Wuhan, arrived in G ansu on 01/18/2020, symptom onset on 01/19/2020, visit clinic on 01/24/2020, hospitalized on 01/24/2020. ## highest: new recovered imported COVID-19 pneumonia patient in Beijing: female, returned to Beijing from Wuhan on 01/08/2020, symptom onset afterwards, recovered on 01/24/2020. new recovered imported COVID-19 pneumonia patient in Beijing: male, returned to Beijing from Wuhan on 01/08/2020, symptom onset afterwards, recovered on 01/25/2020. Second confirmed imported COVID-19 pneumonia patient in Guangxi: male, 46, in contact with individuals from Wuhan before symptom onset. symptom onset on 01/20/2020. Second confirmed imported COVID-19 pneumonia patient in Liaoning: male, 40, works in Wuhan, visit Fushun, Liaonin g on 01/12/2020, symptom onset on 01/14/2020, visit clinic in Fushun Dalian on 01/19/2020. Second confirmed imported COVID-19 pneumonia patient in Sichuan: male, 57, Wuhan resident, visited Sichuan on 01/ 15/2020, symptom onset on 01/16/2020 and hospitalized. ## -----## location n missing distinct ## 1085 0 156 Aichi Prefecture Alappuzha ## lowest : Afghanistan Algeria Amiens ## highest: Yunnan Zabaikalsky Zaragoza Zhejiang Zhuhai ## -----## country n missing distinct ## 1085 0 ## lowest : Afghanistan Algeria Australia Austria Bahrain UK USA ## highest: Thailand UAE Vietnam ## -----## gender n missing distinct 902 183 ## Value female male ## Frequency 382 ## Proportion 0.424 0.576 ## ----n missing distinct Info Mean Gmd . 05 .10 0.999 843 49.48 ## 242 85 20.79 22.0 25.0 35.0 51.0 75.0 ## ## lowest: 0.25 0.5 1 2 4 , highest: 86 87 89 91 96 ## -----## svmptom_onset n missing distinct ## 563 522 70 ## lowest: 01/02/20 01/03/20 01/04/20 01/05/20 01/06/20 ## highest: 2/22/2020 2/23/2020 2/24/2020 2/25/2020 2/26/2020 ## -----## If_onset_approximated n missing distinct Info Sum Mean 525 2 0.123 24 0.04286 0.08219 ## ## -----## hosp_visit_date n missing distinct 507 578 60 ## lowest : 01/01/20 01/03/20 01/05/20 01/06/20 01/08/20 ## highest: 2/24/2020 2/25/2020 2/26/2020 2/27/2020 2/28/2020 ## ------## exposure_start n missing distinct 128 957 39 ## lowest: 01/03/20 01/06/20 01/08/20 01/09/20 01/10/20 ## highest: 2/15/2020 2/17/2020 2/19/2020 2/20/2020 2/21/2020 ## -----## exposure_end n missing distinct 341 744 ## lowest : 01/02/20 01/03/20 01/04/20 01/05/20 01/06/20 ## highest: 2/21/2020 2/22/2020 2/23/2020 2/24/2020 2/25/2020 ## ------## visiting.Wuhan n missing distinct Info Sum Mean 0 2 0.437 192 0.177 0.2916 ## -----## from.Wuhan n missing distinct Info Mean Sum Gmd 1081 4 2 0.37 156 0.1443 0.2472 ## ## death ## n missing distinct 1085 0 14 ## 0 (1022, 0.942), 02/01/20 (1, 0.001), 1 (42, 0.039), 2/13/2020 (1, 0.001), ## 2/14/2020 (1, 0.001), 2/19/2020 (2, 0.002), 2/21/2020 (2, 0.002), 2/22/2020 (1, ## 0.001), 2/23/2020 (4, 0.004), 2/24/2020 (1, 0.001), 2/25/2020 (2, 0.002), ## 2/26/2020 (3, 0.003), 2/27/2020 (2, 0.002), 2/28/2020 (1, 0.001) ## -----## recovered ## n missing distinct 1085 ## lowest : 0 02/02/20 02/04/20 02/05/20 02/06/20 ## highest: 2/24/2020 2/25/2020 2/26/2020 2/27/2020 2/28/2020 ## -----## symptom n missing distinct 270 815 ## ## lowest : chest discomfort chills cold, fever, pneumonia cough cough with sputum ## highest: throat pain, chills throat pain, fever tired vomiting, cough, fever, sore throat vomiting, diarrhea, fever, cough ## source ## n missing distinct 1085 0 ## Aljazeera ## lowest : ABC ABC News Al Arabiya ## highest: Yonnhap News Agency 人民日报 人民日报官方微博 ## -----## link n missing distinct 0 490 ## 1085 ## ## lowest : http://behdasht.gov.ir/news/%DA%A9%D8%B1%D9%88%D9%86%D8%A7+%D9%88%DB%8C%D8%B1%D9%88%D8%B3/199807/%D8% D8%A8%D8%A7+%D8%B9%D9%84%D8%A7%D8%A6%D9%85+%D8%B4%D8%A8%D9%87+%D8%A2%D9%86%D9%81%D9%84%D9%88%D8%A2%D9%86%D8%B2%D 8%A7+%D8%AF%D8%B1+%DA%A9%D8%B4%D9%88%D8%B1+%D8%A8%D8%B3%D8%AA%D8%B1%DB%8C+%D8%B4%D8%AF%D9%86%D8%AF+%D8%A8%D8%B1+% D8%A7%D8%B3%D8%A7%D8%B3+%D8%A2%D8%AE%D8%B1%DB%8C%D9%86+%D9%86%D8%AA%D8%A7%DB%8C%D8%AC+%D8%A2%D8%B2%D9%85%D8%A7%D B%8C%D8%B4+%D9%87%D8%A7+%D8%A7%D8%A8%D8%AA%D9%84%D8%A7%DB%8C+13+%D9%85%D9%88%D8%B1%D8%AF+%D8%AF%DB%8C%DA%AF%D8%B1 + %D8%A8%D9%87 + %D4%A9%D9%88%D9%88%D8%8C%D8%AF19 + %D9%82%D8%B7%D8%B9%DB%8C + %D8%A8%D9%87 + %D9%86%D8%B1 + %D9%85%D8%B1 + %D9%85%D8 + %D9%85%D8%B1 + %D9%85%D8 + %D9%85%D9 + %D9%85%D9 + %D9%85%D9 + %D9%85%D9 + %D9%85%D9 + %D9%95%D9 + %D9%95%D9B%8C+%D8%B1%D8%B3%D8%AF http://english.alarabiya.net/en/News/gulf/2020/02/25/Number-of-Kuwait-coronavirus-cases-r ises-to-eight-KUNA.html http://sxwjw.shaanxi.gov.cn/art/2020/1/27/art_9_67483.html http://wjw.beijing.gov.cn/xwzx_20031/wnxw/202001/t20200121_1620353.html http://wjw.sz.gov.cn/wzx/202001/t20200120_18987787.htm ## highest: https://www3.nhk.or.jp/nhkworld/en/news/20200116_23/ https://www3.nhk.or.jp/nhkworld/en/news/20200124_14/ https://www3.nhk.or.jp/nhkworld/en/news/20200126_31/ https://www3.nhk.or.jp/nhkworld/en/news/20200130_02/ https://www3.nhk.or.jp/nhkworld/en/news/20200131_01/ ## -----## Variables with all observations missing: ## [1] X X.1 X.2 X.3 X.4 X.5 X.6 Add new column to dataframe as current "death" column is inconsistent. After using describe(), I noticed that there are 14 values for "death". To combat, I will create this new column that standardizes "death". updated dataframe Covid_19_rates\$death_updated <- as.integer(Covid_19_rates\$death != 0)</pre> unique(Covid_19_rates\$death_updated) ## [1] 0 1 head(Covid_19_rates, 4) ## id case_in_country reporting.date X NA 1/20/2020 NA ## 1 1 NA 1/20/2020 NA NA 1/21/2020 NA NA 1/21/2020 NA ## 2 2 ## 3 3 ## 4 4 summary ## 1 First confirmed imported COVID-19 pneumonia patient in Shenzhen (from Wuhan): male, 66, shenzheng residence, visited relatives in Wuhan on 12/29/2019, symptoms onset on 01/03/2020, returned to Shenzhen and seek medical car e on 01/04/2020, hospitalized on 01/11/2020, sample sent to China CDC for testing on 01/18/2020, confirmed on 01/ 19/2020. 8 others under medical observation, contact tracing ongoing. ## 2 First confirmed imported COVID-19 pneumonia patient in Shanghai (from Wuhan): female, 56, Wuhan residence, arrive d in Shanghai from Wuhan on 01/12/2020, symptom onset and visited fever clinic on 01/15/2020, laboratory confirme d on 01/20/2020 ## 3 First confirmed imported cases in Zhejiang: patient is male, 46, lives in Wuhan, self-driving from Wuhan to Hangz hou on 01/03/2020, symptom onset 01/04/2020, hospitalized on 01/17/2020, sample deliver to China CDC for testing on 01/20/2020, test positive on 01/21/2020. ## 4 new confirmed imported COVID-19 pneumonia in Tianjin: female, age 60, recently visited Wuhan, visited fever clini c on 01/19/2020 in Tianjin then quarantined immediately. location country gender age symptom_onset If_onset_approximated ## 1 Shenzhen, Guangdong China male 66 01/03/20 Shanghai China female 56 1/15/2020 Shanghai China female 56 1/15/2020 Zhejiang China male 46 01/04/20 ## 2 ## 3 Tianjin China female 60 <NA> ## hosp_visit_date exposure_start exposure_end visiting.Wuhan from.Wuhan death 01/11/20 12/29/2019 01/04/20 1 0 0 ## 1 Official Weibo of Shanghai Municipal Health Commission ## 2 0 ## 3 Health Commission of Zhejiang Province ## 4 人民日报官方微博 ## ## 1 http://wjw.sz.gov.cn/wzx/202001/t20200120_18987787.h tm ## 2 https://www.weibo.com/2372649470/IqogQhgfa?from=page_1001062372649470_profile&wvr=6&mod=weibotime&type=comme nt ## 3 http://www.zjwjw.gov.cn/art/2020/1/21/art_1202101_41786033.ht ml## 4 https://m.weibo.cn/status/446323540126845 7? X.1 X.2 X.3 X.4 X.5 X.6 death_updated ## 1 NA ## 3 NA NA NA NA NA 0 ## 4 NA NA NA NA NA First, I would like to compare the mean age of the people that are alive with covid versus the mean age of those who have contracted covid and have died. I hyothesize that of those who have contracted covid, there is a greater chance someone will die if they are older. alive_subset <- subset(Covid_19_rates, death_updated == 0)</pre> dead_subset <- subset(Covid_19_rates, death_updated == 1)</pre> alive_mean <- mean(alive_subset\$age, na.rm = TRUE)</pre> dead_mean <- mean(dead_subset\$age, na.rm = TRUE)</pre> dead_mean > alive_mean ## [1] TRUE My claim is true as we see that the average age of those who died is greater than the average age of those who are still alive. Next, I want to compare covid related deaths in male versus females to see if there is a difference in mortality rate. By conudcting a t-test, I can determine whether or not this result is signficant. For this t-test, our null hypothesis is that the death rates for male and females are the same and the alternative hypothesis is that they are different. In order to keep all of my test consistent, I will be using a p-value of 0.05. t.test(death_updated ~ gender, data = Covid_19_rates) ## Welch Two Sample t-test ## ## data: death_updated by gender ## t = -3.084, df = 894.06, p-value = 0.002105 ## alternative hypothesis: true difference in means between group female and group male is not equal to 0 ## 95 percent confidence interval: ## -0.07849151 -0.01744083 ## sample estimates: ## mean in group female mean in group male 0.03664921 0.08461538 As we can see from our results, the mean death rate for males is 0.0846 while the mean death rates for females is 0.0366, which at first glance is drastically different. Our t-test also proves so, as we get a resulting p-value of 0.002. Because 0.002 < 0.05, we can comfortably reject our null hypothesis and conlcude that our results our statistically significant and that the death rates do indeed vary between male and females. Another variable that this data set provides it "Visiting Wuhan" expressed as a binary variable. As Wuhan is commonly believed to have orginated in Wuhan China, I thought it would be interesting to test whether or not the death rates in Wuhan were higher than that of the death rates not in Wuhan. Once again, I will be using a two sample t-test with p value of 0.05 to test my hypothesis. The null hypothesis is that death rates inside Wuhan and outside Wuhan are the same but the alternative hypothesis is that the two are different. I begin by making two new subset from from. Wuhan to compare the mean of death rates in these two categories from_wuhan = subset(Covid_19_rates, from.Wuhan == 1) nfrom_wuhan = subset(Covid_19_rates, from.Wuhan == 0) mean(from_wuhan\$death_updated, na.rm = TRUE) ## [1] 0.2179487 mean(nfrom_wuhan\$death_updated, nam.rm = TRUE) ## [1] 0.03135135 t.test(from_wuhan\$death_updated, nfrom_wuhan\$death_updated) ## Welch Two Sample t-test ## ## data: from_wuhan\$death_updated and nfrom_wuhan\$death_updated ## t = 5.5447, df = 164.38, p-value = 1.152e-07 ## alternative hypothesis: true difference in means is not equal to 0 ## 95 percent confidence interval: ## 0.1201495 0.2530453 ## sample estimates: ## mean of x mean of y ## 0.21794872 0.03135135 As we can see from our t-test above, the p-value is extremely small and incomparable to our p-value of 0.05. Thus we can conclude that rejct the null hypothesis and conclude that death rates are in fact different in and out of Wuhan. The second part of this analysis is to measure the effectiveness of vaccines in regards to prevent covid related deaths. usa_vaccines <- read.csv("/Users/dawsontam/Desktop/us_state_vaccinations.csv")</pre> states <- read.csv("/Users/dawsontam/Desktop/states.csv")</pre> describe(usa_vaccines) ## usa_vaccines ## 16 Variables 54628 Observations ## -----## date ## n missing distinct 54628 0 872 ## lowest : 2020-12-20 2020-12-21 2020-12-22 2020-12-23 2020-12-24 ## highest: 2023-05-06 2023-05-07 2023-05-08 2023-05-09 2023-05-10 ## -----## location n missing distinct 54628 0 ## ## lowest : Alabama Alaska American Samoa Arizona Arkansas Washington West Virginia Wisconsin ## highest: Virginia Wyoming ## -----## total_vaccinations n missing distinct Info Mean Gmd .05 .10 32532 22096 30797 1 11851991 18680595 52311 115141 . 50 .75 .90 927731 3199256 7948546 15837354 27258094 ## ## lowest : 416 424 565 ## highest: 674024493 675024615 675442636 676006980 676728782 ## total_distributed n missing distinct Info Mean Gmd 32272 22356 22765 1 14988322 23672190 71460 147460 . 25 .50 .75 . 90 . 95 ## 1188560 4051065 10064899 20590352 35119553 ## lowest : 6000 9175 9200 9575 ## highest: 973224305 978124175 979617855 982117925 984444295 ## -----## people_vaccinated n missing distinct Info Mean .10 1 5909278 9253052 30112 32268 22360 30966 65908 .90 . 50 .75 . 25 . 95 504310 1730820 4023565 7380491 14662996 ## ## lowest : 401 414 416 ## highest: 269965210 269971358 270047396 270129677 270227181 ## ------## people_fully_vaccinated_per_hundred n missing distinct Info Mean Gmd 1 47.33 24.74 3.505 10.360 30651 23977 7992 . 25 . 50 .75 .90 .95 34.240 52.280 62.480 73.240 79.000 ## lowest : 0 0.01 0.02 0.03 0.07 , highest: 91.01 91.04 91.19 91.23 91.26 ## -----## total_vaccinations_per_hundred n missing distinct Info Mean Gmd . 05 .10 30750 23878 15753 1 114.7 66.04 13.27 28.97 .90 . 25 . 50 . 75 . 95 75.02 116.72 154.39 191.26 209.95 ## lowest: 0.17 0.18 0.3 0.41 0.55 , highest: 300.32 300.88 301.05 302.49 302.85 ## -----## people_fully_vaccinated n missing distinct Info Mean . 10 32188 22440 30692 1 4891624 7719620 20069 42128 . 50 . 25 . 75 .90 372328 1349818 3396357 6175936 12527659 1 2 3 4 5 ## lowest : ## highest: 230368815 230485008 230533196 230582699 230637348 ## people_vaccinated_per_hundred n missing distinct 30726 23902 8886 56.57 9.99 18.83 1 26.44 ## . 25 . 50 . 75 .90 . 95 42.59 60.61 72.41 85.82 91.20 ## lowest: 0.24 0.28 0.4 0.54 0.58 , highest: 118.27 118.32 118.46 118.51 118.55 ## -----## distributed_per_hundred n missing distinct Info Mean Gmd . 05 . 10 30730 23898 14839 1 144.7 19.78 .90 . 25 . 50 . 75 . 95 95.63 141.33 196.39 238.34 262.81 ## lowest : 6.14 6.17 6.62 6.63 6.72 , highest: 410.62 411.61 412.14 413.66 414.9 ## -----## daily_vaccinations_raw Mean n missing distinct Info Gmd .10 1 37388 26274 17304 .50 .90 . 25 . 75 . 95 1075 6846 22088 56669 97951 ## lowest : 0 2 1 ## highest: 3965883 3975059 4081959 4575496 4629928 ## -----## daily_vaccinations .10 n missing distinct 54563 22513 204 65 1 25118 42100 .90 . 25 . 50 . 75 . 95 1036 4253 13895 37969 67177 ## lowest : 0 1 2 3 ## highest: 3214893 3330740 3348189 3349306 3384387 ## -----## daily_vaccinations_per_million n missing distinct Info .10 51716 2355 2451 246 2912 8824 1 149 . 25 . 50 . 75 . 90 . 95 627 1498 3172 5794 7611 ## lowest: 0 1 3 4 5, highest: 25957 26252 26597 26960 27652 ## -----## share_doses_used n missing distinct Info Mean Gmd 32272 22356 1 0.7885 0.1192 0.5910 0.6640 932 . 25 . 50 . 75 .90 0.7420 0.7970 0.8520 0.9020 0.9524 ## lowest : 0.043 0.047 0.048 0.058 0.062, highest: 1.27 1.272 1.273 1.275 1.276 ## -----## total_boosters n missing distinct Info Mean 18116 36512 17291 1 2581003 4161231 6387 19299 . 50 .90 . 25 . 75 . 95 145342 605864 1799718 3570672 6393894 ## lowest : 11 14 15 34 ## highest: 190790403 191668184 191961979 192394539 192964188 ## -----## total_boosters_per_hundred .10 n missing distinct Info Mean 17254 37374 5656 2.82 1 25.51 19.9 1.45 . 25 . 50 . 75 .90 . 95 ## lowest: 0 0.01 0.02 0.03 0.04 , highest: 103.01 103.16 103.34 103.66 104.04 ## -----I will now get the mean number of people vaccinated grouped by their state in order to compare the vaccination rate across the state. grouped_us_vaccines <- usa_vaccines %>% group_by(location) %>% summarise(mean_total_vaccinated = mean(people_vaccinated = mean(people_vaccina inated, na.rm = TRUE))grouped_us_vaccines <- grouped_us_vaccines[grouped_us_vaccines\$location %in% states\$State,] grouped_us_vaccines\$state_code <- state.abb[match(grouped_us_vaccines\$location, state.name)]</pre> grouped_us_vaccines_abb <- grouped_us_vaccines[,-c(1)]</pre> grouped_us_vaccines_abb <- grouped_us_vaccines_abb %>% replace(is.na(.), "NY") grouped_us_vaccines_abb <- grouped_us_vaccines_abb[order(grouped_us_vaccines_abb\$state_code),]</pre> colnames(grouped_us_vaccines_abb)[2] = "State" In order to get the death rates, I will have to import another dataset. death_rates <- read.csv("/Users/dawsontam/Desktop/covid_deaths_usafacts.csv")</pre> death_rates_2021_2022 <- death_rates[, -4:-349]</pre> no_na <- death_rates_2021_2022 %>% replace(is.na(.), 0) no_na\$sum_of_row <- rowSums(no_na[, c(4:ncol(no_na))])</pre> cleaned <- no_na[, c("State", "sum_of_row")]</pre> meaned_deaths_per_state <- cleaned %>% group_by(State) %>% summarise(mean_death = mean(sum_of_row)) ordered_mean_deaths <- meaned_deaths_per_state[order(meaned_deaths_per_state\$State),]</pre> ordered_mean_deaths <- ordered_mean_deaths[-8,]</pre> However, it is important to note that different states have different populations so in order to standardize everything I have to import a 2021 - 2022 per state populations dataset. populations <- read.csv("/Users/dawsontam/desktop/NST_EST2022_POP_1_2.csv")</pre> colnames(populations)[1] = "State" new_populations <- populations[, -c(2:5)]</pre> new_populations <- new_populations[order(new_populations\$State),]</pre> new_populations <- new_populations[-1,]</pre> standardized_death <- merge(x = ordered_mean_deaths, y = new_populations, by = "State", all = TRUE) standardized_death <- standardized_death %>% replace(is.na(.), 778597.5) final_standard_death <- standardized_death[-35,]</pre> final_standard_death\$standard_death <- final_standard_death\$mean_death / final_standard_death\$mean_2021_2022 Now that I have my standardized death rates for the 50 states, I can begin a linear regression analysis to see whether not we see a negative correlation between vaccine and death rates. vaccines_vs_death <- $merge(x = final_standard_death, y = grouped_us_vaccines_abb, by = "State", all = TRUE)$ lmvaccine_death <- lm(formula = mean_total_vaccinated ~ standard_death, data = vaccines_vs_death)</pre> summary(lmvaccine_death) ## ## lm(formula = mean_total_vaccinated ~ standard_death, data = vaccines_vs_death) ## ## Residuals: 1Q Median 3Q Min ## -3546478 -2369203 -1114911 1128423 19687419 ## Coefficients: Estimate Std. Error t value Pr(>|t|)## (Intercept) 3945886 728347 5.418 1.92e-06 *** ## standard_death -6144646 5784551 -1.062 0.293 ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## Residual standard error: 4143000 on 48 degrees of freedom ## Multiple R-squared: 0.02297, Adjusted R-squared: 0.002613 ## F-statistic: 1.128 on 1 and 48 DF, p-value: 0.2934 plot(standard_death ~ mean_total_vaccinated, data = vaccines_vs_death, main = "VACCINATION RATE VERSUS DEATH RAT E", xlab = "mean number of vaccinations", ylab = "mean number of covid related deaths") abline(lm(standard_death ~ mean_total_vaccinated, data = vaccines_vs_death)) **VACCINATION RATE VERSUS DEATH RATE** 0 mean number of covid related deaths 5 0 0.4 3 0 0.2 Ö 0 0 0 2.0e+07 0.0e + 005.0e+06 1.0e+07 1.5e+07 mean number of vaccinations As we can see from our linear analysis, we can see that there is indeed a negative correlation between vaccination rates and death rates in the 50 states. Our linear coefficient is 3945886 with a standard error of 728347. Although it is not a severely strong negative linear relationship. It is undeniable that there is a negative correlation between these two factors and thus my orginal hypothesis is corect.

Bangkok Post

li

Covid 19 Analysis