PM-566 Midterm

Jaxon Abercrombie

10/24/2021

Introduction

The ongoing COVID-19 pandemic has come under greater control with the roll-out of COVID-19 vaccines starting back in December 2020. California's governor, Gavin Newsom, speaks regularly about the success that our state has had with controlling the virus through immunization efforts. While California is certainly diverse and varies in demographic composition and physical environment by county, investigating how vaccine uptake has varied by county would be fascinating. Identifying which counties are behind on COVID-19 immunizations would be crucial for intervening and attempting to increase the percent of vaccinated individuals in a more targeted fashion.

Furthermore, toward the beginning of the vaccine roll-out, discourse about which vaccine company people should receive their dose(s) from was frequent. Whether someone received Moderna, Pfizer, or Johnson & Johnson vaccines would dictate whether they received one or two doses and had greater immunity against variants. Because of this, investigating how California counties may differ by vaccine company distribution would also be exciting. Additionally, comparing vaccination rates alongside cases and deaths may reveal changes in vaccination attitudes based on the ebb and flow of COVID-19 morbidity and mortality. For example, if cases for a specific variant grew largely over summer, would it be expected to have more vaccine uptake out of fear?

All that said, the primary question at hand is: *How have COVID-19 vaccination rates varied by county in California since their initial roll-out?* Furthermore, there are two secondary questions that dig deeper into the data sets used: (1) *How do vaccination efforts vary by vaccine company across these counties (Pfizer, Moderna, Johnson & Johnson)?* and (2) *How do trends in cases and deaths potentially affect immunization rates for California as a whole?*

Methods

There were two different sets of data used for this project: one including data for CA counties and their administered vaccine doses and another with data regarding COVID-19 cases and deaths for each CA county. The data about vaccine doses for each county, titled "Vaccines by California County," came from LA city's data site at this link, and it was downloaded by a CSV file. This particular set has data from the start of vaccine roll-out in mid-December 2020 up until mid-October 2021 and includes dose data for each vaccine company per county, county population count, and administration date. Data regarding COVID-related cases and deaths came from the California Department of Health and Human Services at this link, and it was also downloaded as a CSV file. COVID-related cases and deaths, both raw and cumulative, are included for every day since February 2020, and the number of total tests conducted is also included.

Data Wrangling The data was first read and stored in two different variables.

```
vaxCA <- read.csv("CA Vaccine Data.csv")
covidCA <- read.csv("CA Deaths and Cases.csv")
```

Because the sets would be combined by county and date, ensuring consistency in date format and variable name for county was important. After ensuring identical format and variable names, the two data sets were combined. Then, to filter the combined set by when vaccine roll-out began, the date set was altered to contain only dates from December 15, 2020 onward. Additionally, the date variable was altered to not include dates past October 20, 2021 because there was missing data for more recent dates.

Upon browsing the merged data set, it appeared that many rows existed for each county for the dates of August 30, 2021 onward. To fix this, the average was applied for variables for those dates, essentially combining multiple instances into one for the duplicate date-county combinations.

```
covidCA <-
  covidCA %>%
  mutate(county = area)

vaxCA <-
  vaxCA %>%
  mutate(date = as.Date(vaxCA$date, "%m/%d/%Y"))

mergedCA <- merge(vaxCA, covidCA,
                  by = c("date","county"),
                  all.x = TRUE)

mergedCA <-
  mergedCA %>%
  filter(date >= as.Date("2020-12-15")) %>%
  filter(date <= as.Date("2021-10-20")) %>%
  group_by(date, county) %>%
  summarise(across(c(total_doses,cumulative_total_doses,
                     pfizer_doses,cumulative_pfizer_doses,
                     moderna_doses,cumulative_moderna_doses,
                     jj_doses,cumulative_jj_doses,
                     partially_vaccinated,
                     total_partially_vaccinated,fully_vaccinated,
                     cumulative_fully_vaccinated,at_least_one_dose,
                     cumulative_at_least_one_dose,population,
                     cases,cumulative_cases,deaths,
                     cumulative_deaths,total_tests,
                     cumulative_total_tests,positive_tests,
                     cumulative_positive_tests), mean, .groups = date))
```

Additionally, variables were created for later ease with analysis and normalize county data to make it comparable to others. The variables created were transformations of the main variables to be used in analysis:

1. *dose_standard*: the number of cumulative total doses over time standardized by population size to make comparisons
2. *perc_vaccinated* / *perc_partial*: the percent of a county's population that is fully vaccinated / partially vaccinated

```
mergedCA <-
  mergedCA %>%
  mutate(dose_standard = (cumulative_total_doses/population),
         perc_vaccinated = (cumulative_fully_vaccinated/population)*100,
         perc_partial = (cumulative_at_least_one_dose/population)*100)
```

To ensure missing values would not affect our analysis, the number of NAs were summed.

```
mergedNA <- sum(is.na(mergedCA))
```

Fortunately, there were 0 NA values after cleaning and wrangling data.

Exploratory Data Analysis Now that the *mergedCA* data is cleaned and wrangled, it can be explored more. A check on the expected number of observations was performed.

```
expObs <- length(unique(mergedCA$date))*length(unique(mergedCA$county))
actualObs <- nrow(mergedCA)
```

The number of expected observations after wrangling matches the observed number of observations (exp:17980, obs: 17980).

Following this check, summary statistics using the "knitr" package were produced and displayed in the preliminary results section before data visualization and further analysis. Three summary tables were created: one for population data (population count, cases, deaths), one for vaccine status data, and one for vaccine company data.

Data Visualization The main tool used to create visualizations of data was *ggplot2*. The package offered ways to create appealing bar graphs, time series plots, and use *facet_wrap() to* de-clutter plots and focus on data county-by-county.

The visualizations created were:

1. Figure 1: A time series plot depicting how vaccination rates have changed over time by county
2. Figure 2: A bar graph depicting counties by highest percent fully vaccinated, in descending order
3. Figure 3: Pie charts demonstrating the distribution of different vaccine company dose administrations by county
4. Figure 4: A vertically-aligned grid of time series plots (one of total cumulative vaccine doses, one of cases, and one of deaths) to see if trends in mortality/morbidity are related to increased/decreased vaccine hesitancy

Preliminary Results

For some results involving vaccine company differences, it should be acknowledged that Johnson & Johnson requires one dose to be considered fully vaccinated, while it takes two for Moderna and Pfizer. That reality alone may affect statistics involving Johnson & Johnson dose percentages because only one dose would be taken compared to two for other companies' vaccines.

Summary Tables Three summary tables for the data were created to assess minimums, maximums, averages, and standard deviations to ensure no variables were worrisome before continuing with analysis. Variables checked through these tables were those that had not undergone transformation (ex: *fully_vaccinated* rather than *perc_vaccinated* or *cumulative_fully_vaccinated*) to avoid redundancy.

The first table includes data regarding vaccine dose counts for each county, and it also is stratified by vaccine company.

```
mergedCA %>%
  group_by(county) %>%
  summarise(Cases_min = min(cases),
            Cases_mean = mean(cases),
            Cases_max = max(cases),
            Cases_sd = sd(cases),
            Deaths_min = min(deaths),
            Deaths_mean = mean(deaths),
            Deaths_max = max(deaths),
            Deaths_sd = sd(deaths),
            Pop_min = min(population),
            Pop_mean = mean(population),
```

```
            Pop_max = max(population),
            Pop_sd = sd(population)) %>%
  knitr::kable(col.names = c("County",
                              "Min Cases",
                              "Mean Cases",
                              "Max Cases",
                              "SD Cases",
                              "Min Deaths",
                              "Mean Deaths",
                              "Max Deaths",
                              "SD Deaths",
                              "Min Pop",
                              "Mean Pop",
                              "Max Pop",
                              "SD Pop"), digits = 2, "pipe")
```

County	Min Cases	Mean Cases	Max Cases	SD Cases	Min Deaths	Mean Deaths	Max Deaths	SD Deaths	Min Pop	Mean Pop	Max Pop	SD Pop
Alameda	14	240.75	1242	245.10	0	2.86	21	3.85	1685886	1685886	1685886	0
Alpine	0	0.13	3	0.43	0	0.00	0	0.00	1117	1117	1117	0
Amador	0	11.50	121	15.54	0	0.14	2	0.40	38531	38531	38531	0
Butte	2	45.23	191	45.22	0	0.57	5	0.92	217769	217769	217769	0
Calaveras	0	9.71	61	11.37	0	0.18	2	0.45	44289	44289	44289	0
Colusa	0	4.36	29	5.83	0	0.03	1	0.18	22593	22593	22593	0
Contra Costa	10	206.05	931	196.51	0	2.22	15	2.82	1160099	1160099	1160099	0
Del Norte	0	9.35	72	13.28	0	0.12	5	0.46	27558	27558	27558	0
El Dorado	1	34.51	213	35.04	0	0.42	5	0.89	193098	193098	193098	0
Fresno	9	248.90	1369	278.33	0	4.45	23	5.27	1032227	1032227	1032227	0
Glenn	0	6.77	35	7.92	0	0.06	2	0.25	29348	29348	29348	0
Humboldt	2	23.65	91	19.26	0	0.20	3	0.48	134098	134098	134098	0
Imperial	1	38.58	285	48.63	0	1.09	10	2.16	191649	191649	191649	0
Inyo	0	5.19	32	6.27	0	0.07	2	0.26	18453	18453	18453	0
Kern	10	216.49	1271	257.99	0	3.45	18	4.03	927251	927251	927251	0
Kings	0	53.08	271	57.20	0	0.70	5	1.07	156444	156444	156444	0
Lake	0	16.13	79	15.59	0	0.25	4	0.56	64871	64871	64871	0
Lassen	0	9.85	172	19.45	0	0.06	3	0.28	30065	30065	30065	0
Los Angeles	48	2662.55	22267	4386.22	0	56.02	305	82.81	10257557	10257557	10257557	0
Madera	1	42.57	314	49.15	0	0.48	5	0.86	160089	160089	160089	0
Marin	0	26.92	152	27.66	0	0.32	5	0.68	260800	260800	260800	0
Mariposa	0	3.20	30	4.47	0	0.01	1	0.08	17795	17795	17795	0
Mendocino	0	16.58	69	15.70	0	0.19	3	0.49	88439	88439	88439	0
Merced	2	79.95	470	85.18	0	1.17	8	1.39	287420	287420	287420	0
Modoc	0	0.98	16	1.86	0	0.02	2	0.15	9475	9475	9475	0
Mono	0	2.89	31	4.30	0	0.00	0	0.00	13961	13961	13961	0
Monterey	1	80.94	833	145.72	0	1.31	14	2.43	448732	448732	448732	0
Napa	0	25.67	141	27.73	0	0.24	4	0.58	139652	139652	139652	0
Nevada	0	21.07	87	19.42	0	0.16	3	0.50	98710	98710	98710	0
Orange	18	575.54	4385	889.23	0	11.60	69	17.71	3228519	3228519	3228519	0
Placer	5	77.87	341	68.79	0	0.99	8	1.43	400434	400434	400434	0

County	Min Cases	Mean Cases	Max Cases	SD Cases	Min Deaths	Mean Deaths	Max Deaths	SD Deaths	Min Pop	Mean Pop	Max Pop	SD Pop
Plumas	0	3.99	29	5.52	0	0.02	1	0.13	18997	18997	18997	0
Riverside	20	684.69	4839	1034.79	0	10.18	59	14.85	2468145	2468145	2468145	0
Sacramento	22	322.87	1228	268.13	0	4.48	19	4.29	1567975	1567975	1567975	0
San Benito	0	14.42	119	21.33	0	0.15	3	0.41	64022	64022	64022	0
San Bernardino	23	618.63	5284	1035.73	0	11.97	71	18.17	2217398	2217398	2217398	0
San Diego	22	782.72	5255	972.35	0	9.32	58	13.13	3370418	3370418	3370418	0
San Francisco	5	101.48	498	97.85	0	1.48	10	2.21	892280	892280	892280	0
San Joaquin	16	192.60	1101	206.34	0	3.45	20	4.03	782545	782545	782545	0
San Luis Obispo	1	64.32	486	81.80	0	0.86	7	1.40	278862	278862	278862	0
San Mateo	3	102.26	592	119.64	0	1.14	12	2.23	778001	778001	778001	0
Santa Barbara	1	92.05	717	114.50	0	1.21	12	2.03	456373	456373	456373	0
Santa Clara	8	284.55	1754	353.45	0	4.27	33	6.65	1967585	1967585	1967585	0
Santa Cruz	0	40.98	300	57.13	0	0.47	7	1.02	273999	273999	273999	0
Shasta	2	42.27	178	39.80	0	0.78	7	1.16	177925	177925	177925	0
Sierra	0	0.47	7	1.00	0	0.00	0	0.00	3115	3115	3115	0
Siskiyou	0	7.52	45	7.43	0	0.13	2	0.40	43956	43956	43956	0
Solano	1	95.37	574	104.37	0	0.70	5	1.13	444255	444255	444255	0
Sonoma	2	77.41	390	76.65	0	0.75	6	1.19	496668	496668	496668	0
Stanislaus	12	151.91	687	135.36	0	1.93	11	2.33	562303	562303	562303	0
Sutter	0	26.13	110	25.24	0	0.42	5	0.81	105747	105747	105747	0
Tehama	0	17.97	87	19.26	0	0.25	3	0.53	65885	65885	65885	0
Trinity	0	1.35	14	2.22	0	0.01	1	0.10	13354	13354	13354	0
Tulare	2	118.89	682	140.96	0	1.86	12	2.70	484423	484423	484423	0
Tuolumne	0	15.28	311	23.52	0	0.20	3	0.51	52351	52351	52351	0
Ventura	3	211.05	1824	320.80	0	2.97	21	4.47	852747	852747	852747	0
Yolo	1	39.02	214	39.81	0	0.42	5	0.85	223612	223612	223612	0
Yuba	0	21.62	94	20.21	0	0.15	3	0.43	79290	79290	79290	0

Based on the above table, it is evident that more populous counties experience more cases and deaths overall, which is expected and likely relies on population density to happen. Because this data has not yet spanned an entire year, population counts have not updated in the data. Essentially, the population count has remained the same for data despite time elapsing, people being born, and people passing away. The minimum values of the chosen variables do not reach into negative values, and the maximum values are not unusually high, which is good. Data in this table was validated by an external source by comparing some counties' maximum cases and deaths to the source's dashboard.

The second table includes data about vaccination status by county, split up by whether individuals were partially or fully vaccinated.

```r
mergedCA %>%
  group_by(county) %>%
  summarise(Partially_min = min(partially_vaccinated),
            Partially_mean = mean(partially_vaccinated),
            Partially_max = max(partially_vaccinated),
            Partially_sd = sd(partially_vaccinated),
            Fully_min = min(fully_vaccinated),
            Fully_mean = mean(fully_vaccinated),
            Fully_max = max(fully_vaccinated),
            Fully_sd = sd(fully_vaccinated),) %>%
  knitr::kable(col.names = c("County",
                             "Min Partial Vax",
                             "Mean Partial Vax",
                             "Max Partial Vax",
                             "SD Partial Vax",
                             "Min Full Vax",
                             "Mean Full Vax",
                             "Max Full Vax",
                             "SD Full Vax"), digits = 2, "pipe")
```

County	Min Partial Vax	Mean Partial Vax	Max Partial Vax	SD Partial Vax	Min Full Vax	Mean Full Vax	Max Full Vax	SD Full Vax
Alameda	41	3708.75	17155	3667.65	0	3778.43	15691	3945.11
Alpine	0	2.72	72	7.98	0	2.23	86	8.04
Amador	0	70.00	951	104.70	0	55.85	470	82.65
Butte	0	334.70	2420	411.74	0	325.81	2617	416.36
Calaveras	0	75.92	619	105.69	0	67.12	589	100.60
Colusa	0	35.31	336	48.25	0	33.69	308	48.40
Contra Costa	20	2660.88	12144	2695.96	0	2659.40	12122	2839.15
Del Norte	0	37.26	310	51.21	0	36.28	302	50.47
El Dorado	0	342.61	2013	340.11	0	333.65	1761	349.83
Fresno	3	1731.72	9066	1641.78	0	1622.02	8196	1668.54
Glenn	0	42.76	416	63.82	0	43.17	511	69.59
Humboldt	0	250.19	1594	300.72	0	249.40	1872	314.99
Imperial	0	427.25	3190	450.68	0	393.46	2733	444.43
Inyo	0	30.81	504	53.04	0	29.70	505	52.06
Kern	1	1333.04	4657	1062.78	0	1261.79	5992	1125.78
Kings	0	189.38	1573	191.59	0	177.59	1258	194.04
Lake	0	103.79	575	120.72	0	99.98	872	142.08
Lassen	0	22.46	1057	64.27	0	23.46	921	57.63
Los Angeles	373	20846.27	74576	18108.62	0	19970.44	76752	18681.70
Madera	0	239.18	1123	231.44	0	225.38	1184	240.04
Marin	0	650.15	3301	756.46	0	644.96	3363	778.23
Mariposa	0	27.12	372	52.26	0	19.44	397	42.39
Mendocino	0	176.58	1878	268.31	0	165.98	1859	263.54
Merced	0	433.54	2351	417.51	0	357.70	2026	364.14
Modoc	0	10.35	173	24.75	0	10.87	186	26.97
Mono	0	29.18	819	89.39	0	27.69	824	88.91
Monterey	0	878.02	4963	958.27	0	849.64	5373	981.84
Napa	0	322.96	2312	402.93	0	304.90	1966	372.44

County	Min Partial Vax	Mean Partial Vax	Max Partial Vax	SD Partial Vax	Min Full Vax	Mean Full Vax	Max Full Vax	SD Full Vax
Nevada	0	188.92	1004	211.56	0	178.45	1079	210.44
Orange	18	6633.15	25898	5887.00	0	6444.60	23856	6117.61
Placer	1	762.54	2903	681.22	0	740.55	3176	700.27
Plumas	0	29.14	692	67.12	0	30.24	665	71.94
Riverside	8	4256.84	16784	3648.29	0	3993.04	15375	3662.45
Sacramento	5	2965.57	11618	2411.02	0	2845.65	10954	2448.18
San Benito	0	124.06	612	124.01	0	118.34	700	130.46
San Bernardino	13	3557.63	12792	2811.31	0	3366.73	11907	2879.26
San Diego	154	7348.08	24432	6028.42	0	6990.02	26749	6273.51
San Francisco	5	2102.60	9775	2273.60	0	2106.80	9971	2403.52
San Joaquin	19	1390.77	4954	1089.58	0	1214.32	5812	1067.10
San Luis Obispo	1	532.08	2743	607.93	0	514.71	3246	623.91
San Mateo	25	1819.36	7592	1826.57	0	1800.65	7280	1902.03
Santa Barbara	0	888.30	4951	1005.84	0	862.67	4865	1037.02
Santa Clara	67	4613.35	34502	5363.89	0	4656.38	31582	5435.47
Santa Cruz	2	598.43	3028	638.40	0	568.83	2923	626.68
Shasta	0	241.30	1996	278.40	0	227.89	1986	271.82
Sierra	0	5.01	133	14.03	0	4.95	131	13.95
Siskiyou	0	64.40	1340	119.70	0	59.71	1204	111.23
Solano	4	888.29	5038	867.09	0	801.01	4895	831.11
Sonoma	4	1115.95	4519	1163.69	0	1086.50	5388	1193.46
Stanislaus	0	994.70	4222	866.39	0	849.25	4503	810.82
Sutter	0	165.01	954	167.68	0	159.06	1393	197.10
Tehama	0	78.54	544	97.88	0	75.68	594	98.51
Trinity	0	16.96	327	37.85	0	16.74	430	40.20
Tulare	0	722.46	3525	683.69	0	661.54	3855	695.52
Tuolumne	0	89.62	1253	165.70	0	80.70	1203	161.74
Ventura	5	1751.57	8294	1660.59	0	1702.70	6841	1698.75
Yolo	1	446.55	2276	457.24	0	426.10	2289	449.68
Yuba	0	103.08	480	86.94	0	102.80	592	106.66

Based on the above table, the number progression from partially vaccinated to fully vaccinated makes sense; there are overall less fully vaccinated people than partially vaccinated people. In other words, to be fully vaccinated means you reached partial vaccination at one point. Additionally, partial vaccination has minimum values that extend beyond 0 while full vaccination does not. This reflects how someone cannot be fully vaccinated immediately and needs a first dose to reach that status eventually. No values appear abnormal in this table as either, fortunately.

The third table includes data regarding vaccine dose counts and also is stratified by vaccine company.

```r
mergedCA %>%
  group_by(county) %>%
  summarise(Doses_min = min(total_doses),
            Doses_mean = mean(total_doses),
            Doses_max = max(total_doses),
            Doses_sd = sd(total_doses),
            JJ_min = min(jj_doses),
            JJ_mean= mean(jj_doses),
            JJ_max = max(jj_doses),
            JJ_sd = sd(jj_doses),
            Mod_min = min(moderna_doses),
            Mod_mean= mean(moderna_doses),
            Mod_max = max(moderna_doses),
            Mod_sd = sd(moderna_doses),
            Pfi_min = min(pfizer_doses),
            Pfi_mean= mean(pfizer_doses),
            Pfi_max = max(pfizer_doses),
            Pfi_sd = sd(pfizer_doses)) %>%
  knitr::kable(col.names = c("County",
                             "Min Doses",
                             "Mean Doses",
                             "Max Doses",
                             "SD Doses",
                             "Min JJ Doses",
                             "Mean JJ Doses",
                             "Max JJ Doses",
                             "SD JJ Doses",
                             "Min Mod Doses",
                             "Mean Mod Doses",
                             "Max Mod Doses",
                             "SD Mod Doses",
                             "Min Pfi Doses",
                             "Mean Pfi Doses",
                             "Max Pfi Doses",
                             "SD Pfi Doses"), digits = 2, "pipe")
```

County	Min Doses	Mean Doses	Max Doses	SD Doses	Min JJ Doses	Mean JJ Doses	Max JJ Doses	SD JJ Doses	Min Mod Doses	Mean Mod Doses	Max Mod Doses	SD Mod Doses	Min Pfi Doses	Mean Pfi Doses	Max Pfi Doses	SD Pfi Doses
Alameda	41	7797.52	35711	6764.31	0	367.97	8037	1050.28	0	2340.10	10209	2463.82	41	5089.51	15506.00	4085.26
Alpine	0	5.09	117	13.19	0	0.02	1	0.15	0	4.92	116	13.17	0	0.14	2.00	0.41
Amador	0	129.16	986	157.75	0	5.00	254	16.31	0	85.47	906	141.60	0	38.69	233.00	33.27
Butte	0	680.79	3374	691.56	0	27.23	482	58.55	0	304.42	2664	450.92	0	349.14	2729.00	352.83
Calaveras	0	148.23	959	173.60	0	4.68	75	7.83	0	76.92	748	122.31	0	66.64	866.00	103.71
Colusa	0	69.87	445	83.38	0	2.65	71	7.93	0	45.79	393	71.17	0	21.43	141.00	18.62
Contra Costa	21	5556.41	29247	4952.60	0	177.75	4629	453.28	0	1545.27	7050	1769.50	20	3833.40	14804.00	3328.14
Del Norte	0	76.49	500	87.70	0	4.17	77	9.28	0	41.76	351	62.57	0	30.55	175.00	34.73
El Dorado	4	704.90	3264	614.43	0	27.12	458	48.82	0	320.72	1394	334.86	0	357.06	1875.00	328.95
Fresno	3	3456.07	12603	2861.69	0	118.73	1426	231.58	0	1381.40	7193	1589.35	1	1955.95	6828.00	1294.03

County	Min Doses	Mean Doses	Max Doses	SD Doses	Min JJ Doses	Mean JJ Doses	Max JJ Doses	SD JJ Doses	Min Mod Doses	Mean Mod Doses	Max Mod Doses	SD Mod Doses	Min Pfi Doses	Mean Pfi Doses	Max Pfi Doses	SD Pfi Doses
Glenn	0	87.56	672	109.55	0	4.54	255	16.86	0	42.20	457	76.12	0	40.83	349.00	48.17
Humboldt	0	520.38	2311	504.53	0	25.76	470	49.52	0	218.57	2150	297.42	0	276.06	1581.00	282.52
Imperial	0	837.15	5050	814.96	0	51.40	1637	150.39	0	334.03	2375	391.16	0	451.72	3547.00	454.25
Inyo	0	62.91	568	81.36	0	1.84	128	8.45	0	35.05	245	52.02	0	26.02	536.00	64.84
Kern	1	2650.67	10386	2012.37	0	112.73	1382	231.50	0	1083.16	4220	1022.68	1	1454.79	5494.00	1015.12
Kings	0	373.58	1654	324.43	0	17.67	389	44.27	0	167.55	1558	212.64	0	188.37	1088.00	144.56
Lake	0	208.43	1122	214.92	0	9.50	476	33.64	0	128.11	1005	170.28	0	70.82	839.00	91.24
Lassen	0	46.83	1077	89.63	0	4.21	78	8.59	0	34.28	1065	87.89	0	8.34	86.00	12.71
Los Angeles	410	42190.23	157449	33594.64	0	1594.79	24711	3589.47	0	16347.60	62780	16603.92	246	24247.77	71114.00	16040.87
Madera	0	475.73	1969	408.95	0	18.37	358	39.28	0	196.26	1210	228.22	0	261.10	1338.00	226.24
Marin	0	1363.59	5901	1350.98	0	42.57	742	94.14	0	441.83	2582	576.30	0	879.20	4176.00	836.01
Mariposa	0	47.41	499	79.17	0	2.00	39	4.64	0	33.90	488	75.89	0	11.52	121.00	14.11
Mendocino	0	359.00	2248	430.02	0	11.95	247	24.84	0	144.55	1958	262.30	0	202.49	1805.00	275.20
Merced	0	817.73	3630	711.77	0	31.98	833	83.84	0	284.36	1460	296.15	0	501.39	2557.00	429.14
Modoc	0	21.85	209	40.89	0	1.47	109	6.47	0	16.36	198	38.21	0	4.02	99.00	9.00
Mono	0	59.41	874	130.78	0	1.10	29	2.88	0	23.60	527	75.50	0	34.71	604.00	93.36
Monterey	0	1774.78	10353	1703.66	0	73.40	2237	188.12	0	807.61	4517	958.01	0	893.77	4942.00	837.65
Napa	0	653.25	3105	659.75	0	18.87	380	43.88	0	277.07	1885	378.92	0	357.30	2244.00	357.75
Nevada	0	383.60	1592	371.22	0	11.02	442	29.18	0	159.46	887	184.46	0	213.12	931.00	210.87
Orange	18	13581.43	52936	10543.02	0	417.16	5607	844.82	0	5122.90	19820	5081.05	17	8041.27	25101.00	5429.81
Placer	1	1574.84	7768	1208.37	0	43.23	781	80.42	0	587.71	2265	582.97	1	943.93	3033.00	672.01
Plumas	0	60.38	716	104.69	0	3.93	286	19.65	0	46.26	711	100.09	0	10.19	141.00	14.94
Riverside	8	8516.82	37774	6503.00	0	242.49	4771	497.53	0	3128.41	12615	3190.01	8	5145.96	15628.00	3368.91
Sacramento	5	6038.01	18514	4415.30	0	172.18	3004	314.27	0	2305.39	9727	2199.56	4	3560.43	10514.00	2327.58
San Benito	0	248.48	1038	225.95	0	10.81	258	28.13	0	93.17	597	109.10	0	144.50	748.00	146.55
San Bernardino	13	7101.39	22233	5218.34	0	215.44	3754	442.18	0	2470.18	9191	2454.37	13	4415.77	11655.00	2753.18
San Diego	175	14661.46	50105	10778.11	0	548.56	5917	899.63	0	5815.70	21632	5590.41	175	8297.64	26508.00	5561.53
San Francisco	9	4402.21	37069	4186.70	0	179.02	4491	464.64	0	1328.51	6503	1465.42	9	2894.70	10132.00	2669.37
San Joaquin	19	2676.80	9241	1995.74	0	84.20	2173	242.29	0	1004.95	3905	942.70	16	1587.65	5027.00	1071.26
San Luis Obispo	1	1093.14	4676	1119.09	0	40.48	777	88.42	0	465.01	2574	538.52	0	587.62	2514.00	608.83
San Mateo	25	3784.01	23379	3336.20	0	134.31	2638	366.66	0	1245.22	7482	1434.98	25	2404.49	8468.00	2001.24
Santa Barbara	0	1810.88	6709	1753.30	0	76.02	1731	184.93	0	642.53	3160	683.30	0	1092.31	5537.00	1130.13
Santa Clara	67	9616.40	45081	9416.19	0	346.31	7737	935.22	0	2937.50	15964	3236.22	20	6332.59	34818.00	6445.78

County	Min Doses	Mean Doses	Max Doses	SD Doses	Min JJ Doses	Mean JJ Doses	Max JJ Doses	SD JJ Doses	Min Mod Doses	Mean Mod Doses	Max Mod Doses	SD Mod Doses	Min Pfi Doses	Mean Pfi Doses	Max Pfi Doses	SD Pfi Doses
Santa Cruz	2	1202.80	4826	1125.48	0	31.06	482	62.53	0	559.76	2674	629.43	2	611.98	3049.00	601.47
Shasta	0	480.95	3064	459.43	0	24.93	861	61.15	0	226.98	2329	278.09	0	229.03	2489.00	283.38
Sierra	0	10.11	193	25.11	0	0.19	8	0.68	0	8.59	192	25.16	0	1.33	12.83	2.27
Siskiyou	0	128.49	1700	190.73	0	3.73	64	8.82	0	68.21	674	97.49	0	56.56	1198.00	138.75
Solano	4	1749.16	3688	1475.05	0	47.73	874	96.97	0	698.68	4878	941.25	1	1002.72	5117.00	737.87
Sonoma	4	2294.87	7990	2112.92	0	71.88	1378	150.90	0	898.14	4062	1105.39	1	1324.86	4768.00	1081.48
Stanislaus	0	1895.86	8554	1512.97	0	61.80	1981	184.65	0	694.88	3821	737.46	0	1139.20	4251.00	879.22
Sutter	0	331.45	1700	300.83	0	14.69	633	54.88	0	168.77	1161	205.22	0	147.99	885.00	135.24
Tehama	0	156.42	779	155.94	0	6.11	73	10.23	0	89.27	708	115.01	0	61.04	492.00	83.57
Trinity	0	34.11	463	61.48	0	2.31	110	8.61	0	23.76	389	58.00	0	8.04	38.00	6.68
Tulare	0	1420.35	7437	1147.32	0	38.81	605	81.27	0	507.26	2333	541.17	0	874.30	3450.00	666.22
Tuolumne	0	175.55	1465	255.26	0	4.23	47	6.70	0	117.72	1406	223.42	0	53.59	639.00	97.85
Ventura	5	3568.01	11898	3059.02	0	123.56	2170	274.63	0	1449.72	6923	1637.01	5	1994.81	5970.00	1428.03
Yolo	1	905.61	3876	814.17	0	26.42	565	63.55	0	411.30	2229	468.59	1	467.89	1995.00	386.56
Yuba	0	209.50	795	161.47	0	12.75	491	42.08	0	109.89	614	113.31	0	86.85	410.00	62.59

This table provides the most unique information to me compared to the other two. Again, minimums and maximums for these values do not appear to deviate from normal, expected values. Because the two data sets used in this analysis were downloaded from reputable, government agencies, the data quality is reliable. Already there are clear differences in vaccine company means, with Pfizer most often administered, Moderna second, and Johnson & Johnson third. Visualizing this through pie charts will be interesting to see how the total doses are shared by company for each county.

Figures Figure 1: Vaccination Rates Over Time by County

The following plots shows how administered vaccine doses have accumulated since December 2020. Each county has its own plot, and the y-variable was standardized earlier to make counties' data comparable.

```
mergedCA %>%
  ggplot(data, mapping = aes(x = date, y = dose_standard,
                             color = county)) +
  geom_line() +
  xlab("Doses per Person") +
  facet_wrap(vars(county), ncol = 5) +
  theme(legend.position = "none", strip.background = element_blank(),
        strip.text = element_text(size = rel(0.8), margin = margin()),
        panel.spacing = unit(3, "pt")) +
  labs(x = "Date", y = "Cumulative Total Doses (Standardized by Population Size)")
```

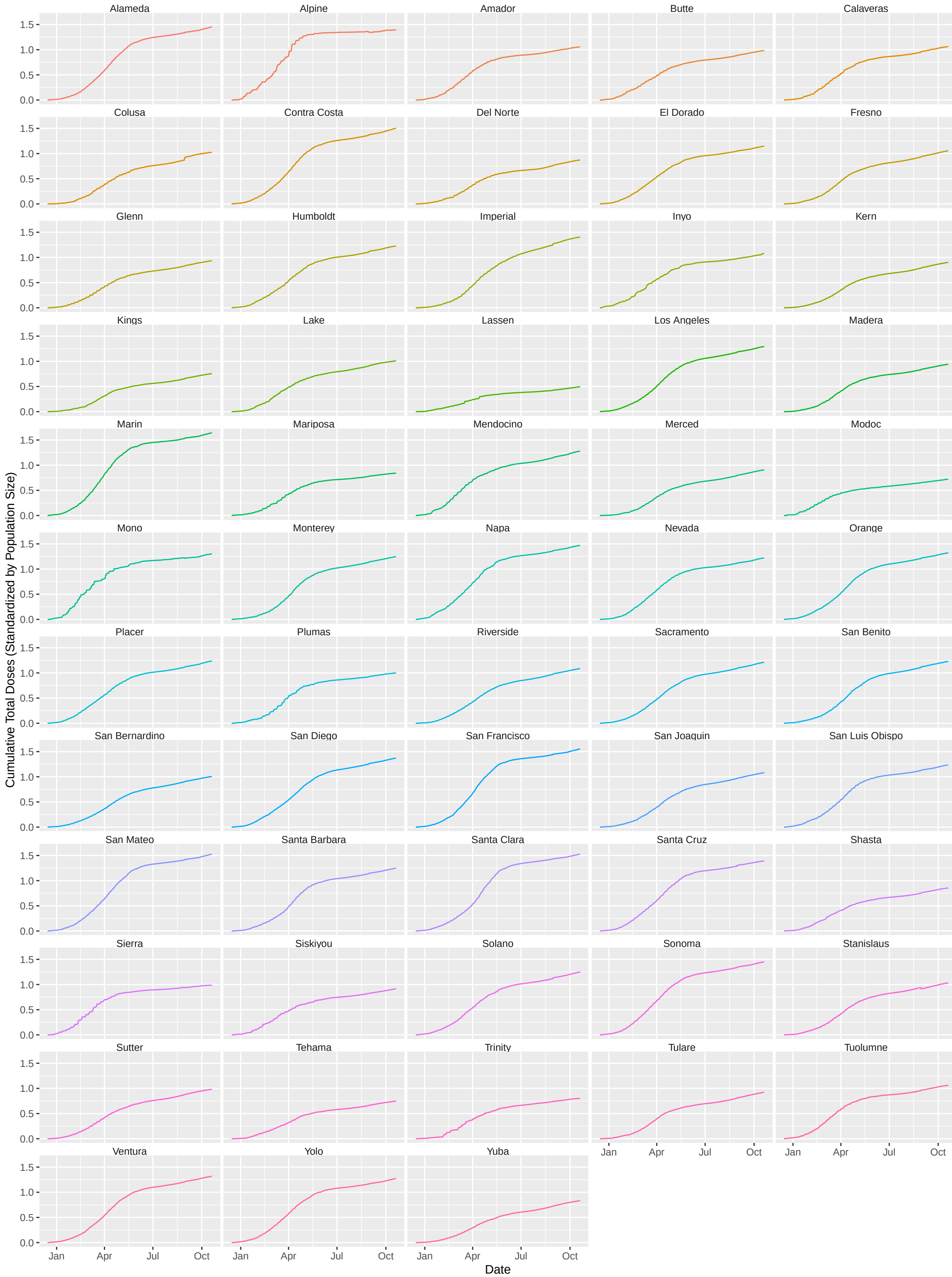


Figure 2: Percent Fully Vaccinated by County

The following figure shows the percent of individuals fully vaccinated by county in descending order.

```
mergedCA <-
  mergedCA %>%
  group_by(county) %>%
  mutate(maxPerc = max(perc_vaccinated)/310)

mergedCA %>%
  ggplot(aes(y = reorder(county, maxPerc), x = maxPerc)) +
  geom_bar(stat = "identity") +
  labs(y = "County", x = "Percent Fully Vaccinated by 10/20/2021") +
  theme(legend.position = "none")
```

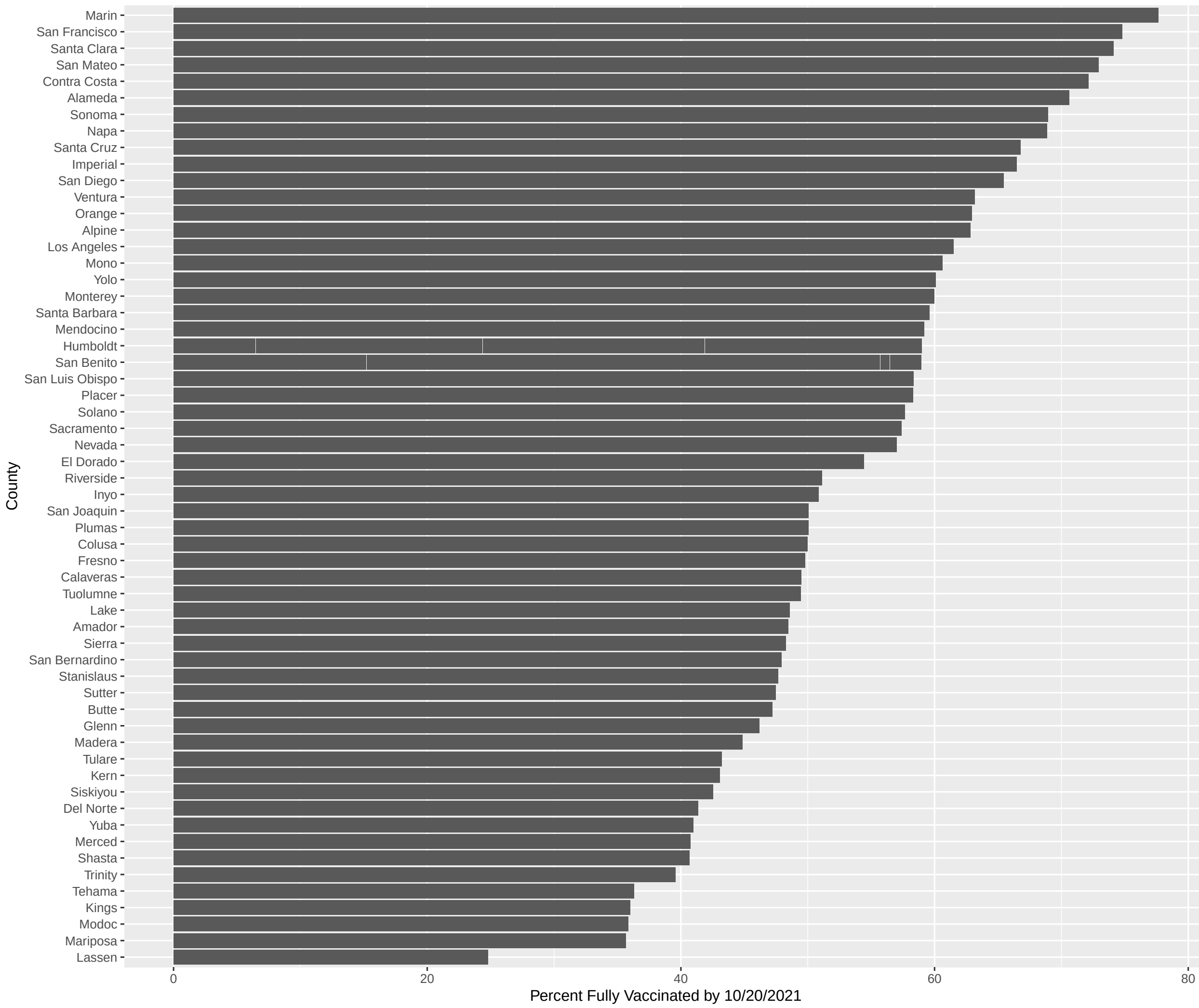


Figure 3: Pie Chart of Vaccines Administered Based on Company across Counties

The following pie charts demonstrate how the administered doses vary by vaccine company for each county.

```
companyVax <-
  mergedCA %>%
  group_by(county) %>%
  summarise(maxJJ =
              (max(cumulative_jj_doses)/max(cumulative_total_doses))*100,
            maxMod =
```

```r
              (max(cumulative_moderna_doses)/max(cumulative_total_doses))*100,
            maxPfi =
              (max(cumulative_pfizer_doses)/max(cumulative_total_doses)*100))

county <- companyVax$county
JJ <- companyVax$maxJJ
Moderna <- companyVax$maxMod
Pfizer <- companyVax$maxPfi
df <- data.frame(county, JJ, Moderna, Pfizer)

require(tidyr)
companyVax <- gather(df, variable,value, -county)

companyVax %>%
  ggplot(aes(x="", y=value, fill=variable)) +
  geom_bar(stat="identity") +
  coord_polar("y", start=0) +
  facet_wrap(vars(county), ncol = 8) +
  labs(x = "", y = "", legend = "Vaccine Company") +
  theme(axis.text.x=element_blank()) +
  scale_fill_brewer("Blues")
```

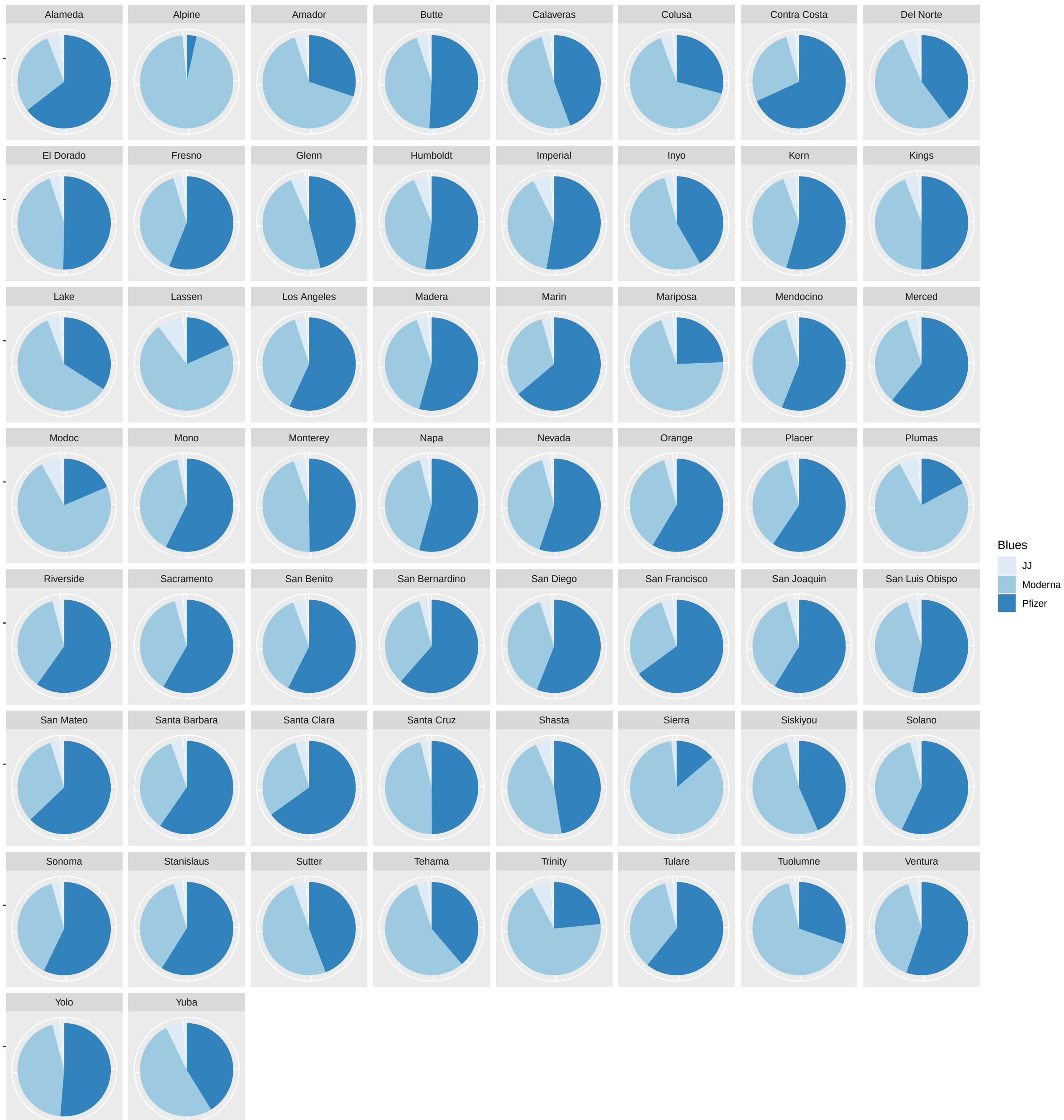
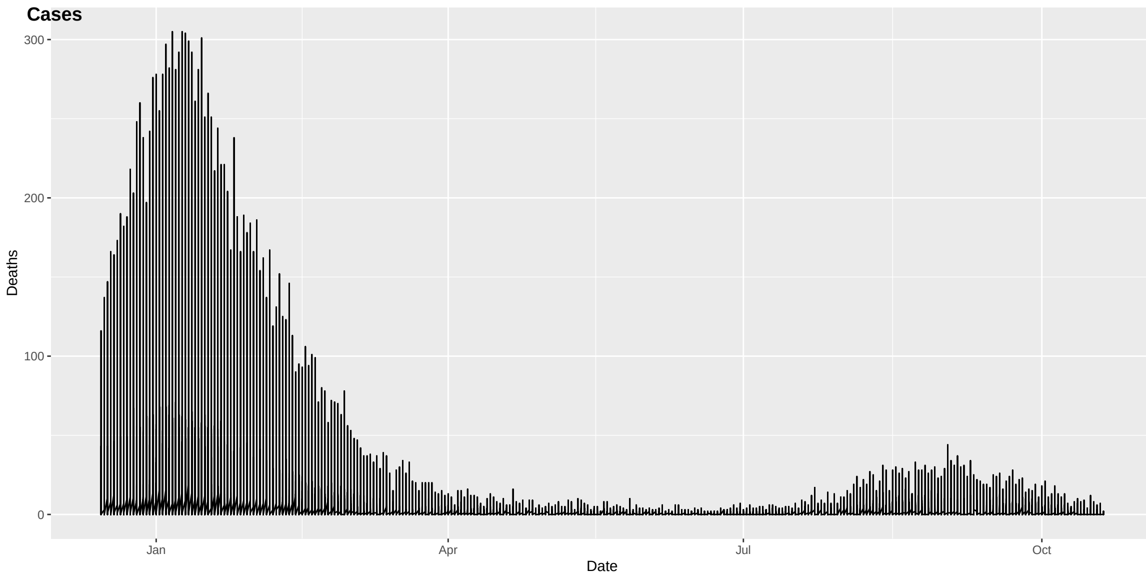


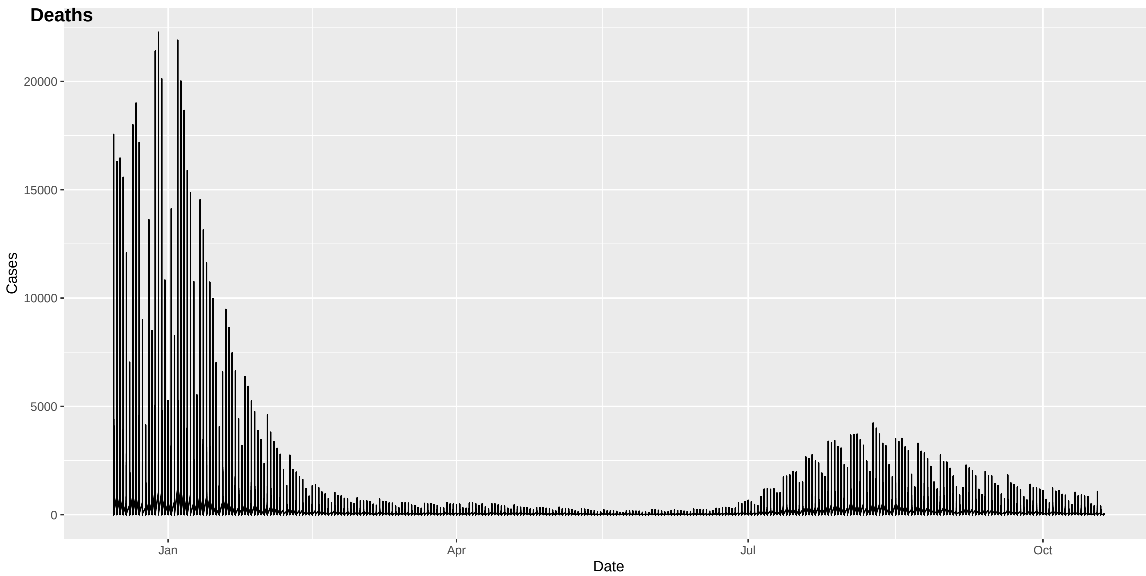
Figure 4: Comparing Trends in Vaccination to Trends in COVID-19 Cases and Deaths (All of CA)

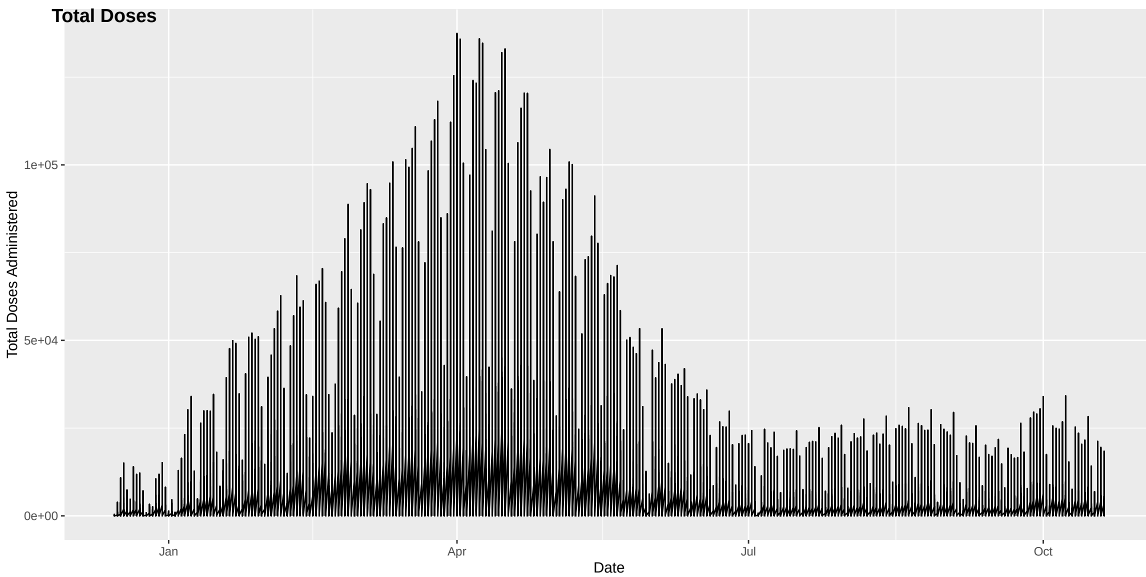
The following three figures are aligned vertically so that trends can be acknowledged based on date. The first plot depicts cases since the start of vaccine roll-out. The second plot depicts deaths since the start of vaccine roll-out. The third plot depicts the total number of doses administered per day since the start of vaccine roll-out. Focus will be placed on the month of July and onward.

```
pCases <-
  mergedCA %>%
  ggplot(mapping = aes(x = date,
                       y = cases)) +
  geom_line() +
  xlab("") +
```

```r
  theme(legend.position = "none", strip.background = element_blank(),
        strip.text = element_text(size = rel(0.8), margin = margin()),
        panel.spacing = unit(3, "pt")) +
  labs(x = "Date", y = "Cases")

pDeaths <-
  mergedCA %>%
  ggplot(mapping = aes(x = date,
                       y = deaths)) +
  geom_line() +
  xlab("") +
  theme(legend.position = "none", strip.background = element_blank(),
        strip.text = element_text(size = rel(0.8), margin = margin()),
        panel.spacing = unit(3, "pt")) +
  labs(x = "Date", y = "Deaths")


pDoses <-
  mergedCA %>%
  ggplot(mapping = aes(x = date,
                       y = total_doses)) +
  geom_line() +
  xlab("") +
  theme(legend.position = "none", strip.background = element_blank(),
        strip.text = element_text(size = rel(0.8), margin = margin()),
        panel.spacing = unit(3, "pt")) +
  labs(x = "Date", y = "Total Doses Administered")


plot_grid(pDeaths, pCases, pDoses,
          ncol = 1,
          labels = c("Cases", "Deaths", "Total Doses"))
```


Cases
Deaths
300
200
100
0
Jan
Apr
Jul
Oct
Date


Deaths
Cases
20000
15000
10000
5000
0
Jan
Apr
Jul
Oct
Date


Total Doses
Total Doses Administered
1e+05
5e+04
0e+00
Jan
Apr
Jul
Oct
Date

Conclusion

Conclusions to the primary question and two secondary questions of this data project were found.

In regard to vaccination differences across California counties, it is evident that:

- All counties follow an S-like curve when looking at cumulative vaccine doses, though steepness differs (Figure 1). The S-shape explains a surge in vaccination rates, with cumulative dose counts plateauing at extremes. This is explainable by eligibility, since at the start of vaccine roll-out (Jan-March 2021), not many people could get vaccinated—just the elderly and healthcare officials. Now, past July 2021, we are seeing a plateau because most individuals who wanted vaccine doses received them.
  - After around April, when eligibility to get a vaccine continued to widen, a greater amount of individuals were able to receive their vaccine and did so. Notably steep increases in vaccination appear to take place in counties like Napa, Santa Clara, Alpine, Alameda, and San Francisco, while significantly less steep increases appear in counties like Yuba, Kern, and Modoc.
    * This generally seems to portray that more urban or suburban counties have greater rates of fully vaccinated individuals than those that have rural settings.
  - Regardless of trends by county, there have been general increases in doses being administered each day despite the surge of getting vaccinated slowing down.
- The majority of counties in Northern California (Marin, San Francisco, Santa Clara) have a greater proportion of fully vaccination individuals than countries in Central (Tulare, Fresno) and Southern California (San Bernardino, Riverside) based on Figure 2.
  - The lowest percentage of fully vaccinated individuals belongs to Lassen County, with about 25% of individuals fully vaccinated. On the other hand, Marin County has the most with about 78%.
  - Roughly 25% of counties have more than 60% of their populations fully vaccinated, and around 50% of counties between 45% and 60% are fully vaccinated. Overall the numbers are low compared to the state's goals.
- Much of these differences could be most attributable to vaccine access and/or political affiliation. Since more rural areas likely face travel-related obstacles to receive vaccines and more conservative areas likely hold anti-vaccine sentiment, it seems reasonable that these trends exist in the data.
  - Of the two, vaccine access may pose a larger threat. This is because many of the counties in the lower half of Figure 2 are the farthest from urban centers and likely have less vaccination services within their counties.

When assessing whether specific vaccine companies were more prevalent in some areas than others, the pie charts produced tell an intriguing story about whether Moderna or Pfizer dominate certain counties.

- The third summary statistic table demonstrates that Pfizer doses are given most often, Moderna second, and Johnson & Johnson third.
- Based on Figure 3, we can see that Moderna and Pfizer essentially take turns with being a county's most popular vaccine. Johnson & Johnson, even when acknowledging that it is a single-dose vaccine, consistently makes up small portions of the charts. Despite Johnson & Johnson being easier with one dose, individuals still appear to receive other companies' vaccines more often.
- It appears that the Pfizer vaccine dominates in the most well-known urban centers like Los Angeles, San Francisco, and San Diego Counties while Moderna is the more common vaccine in rural places like Sierra, Shasta, Plumas, and Lake Counties—just to name a few (Figure 3).
- This difference in vaccine prevalence between urban and rural areas is neat to see. It is possible that this could be attributed to Pfizer being the first vaccine released and the most populous/urban counties having robust vaccination centers to serve their people with Pfizer. Pfizer being the first available vaccine would significantly impact the number of individuals who have it since it had great demand.

Also, since Moderna has longer wait periods between doses and Johnson & Johnson was removed from the market at one point, Pfizer's domination is even more reasonable.

When assessing whether trends in cases and deaths potentially caused trends in vaccine, there is a slight increase in vaccinations when variants were widespread over the summer of 2021 based on Figure 4.

- It appears that the increase in deaths and cases between July and August could have influenced the visible increase in vaccine doses administered during the same time and up until now, October 2021. Because the summer was a great demonstration of how vaccines protect people and prevent severe illness and death, it makes sense that people who originally were skeptical changed their mind.
- It appears that another small vaccination surge happened in October, and any increase in vaccination is beneficial. Especially as we gear up for winter months, increasing the percent of individuals vaccinated is important.
- Now that booster shots are becoming a hot topic, it is unclear whether the dose surge in October could be attributed to booster shots or people receiving their first or second doses.

Overall, this was an extremely exciting project to pursue. Especially as the motivation of unvaccinated individuals to get vaccinated is more important each day, identifying which counties can improve on vaccination is crucial. The tables and plots produced can also help answer several other COVID-related questions, and they assisted in finding trends related to vaccine companies and state-wide vaccination trends that are likely influenced by cases and deaths. In the future, it would be neat to incorporate data about booster shots, do this analysis on a country-wide scale, analyze how COVID-19 fluctuates this winter, and produce maps of California with the information and even more unique visualizations.