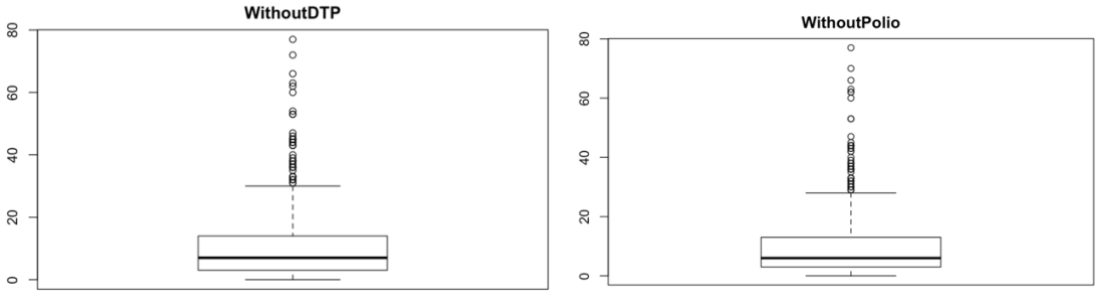
Descriptive Reporting:

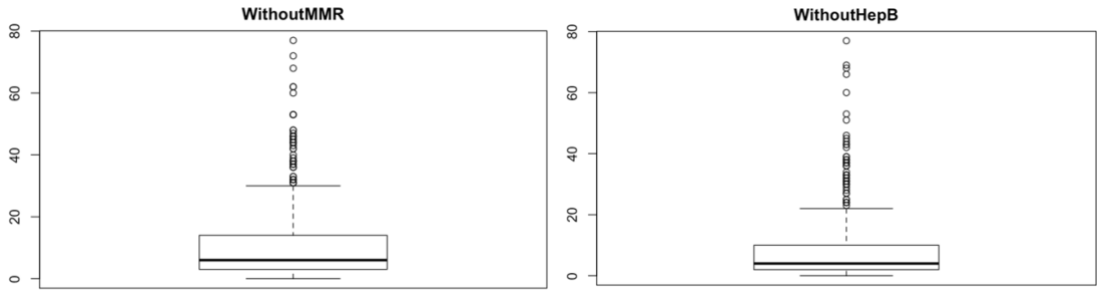
1. This report includes U.S Vaccinations dataset reporting vaccination rates in the U.S. for five common vaccines and districts dataset reflecting California public school districts with specific numbers and percentages for each district. You can see clearly what level of California vaccination rate compared with U.S vaccination. In the conclusion, you can see how to improve vaccination rates and reporting compliance to California school districts.

2. Overview of U.S. Vaccinations
    a. U.S. vaccination rates over time
        i. From the recent (2010-2017) vaccination's rete, the overall rate is around 95%. We can see the vaccinate rate of Diptheria, Tetanus, Pertussis had a little increase, around 1% or 2%, and then became stable. In general, the fluctuation was very small.
        ii. The vaccinate rate of Haemophilus influenza type B vaccine had a 10% increase, and then became stable. The overall rate is around 90%.
        iii. The vaccinate rate of Hepatitis B, Birth Dose had a 20% increase from 50% to 70% between 2010-2013, and then had a 10% decrease between 2013-2017.
        iv. The vaccinate rate of Measles-containing-vaccine first-dose had some fluctuations but finally stabilized at 93%.
        v. The vaccinate rate of Inactivated polio vaccine was very stable around 97% or so.
    b. Notable trend was answered in part a
    c. Mean U.S. vaccination rate from 2010-2017 (%)

DTP1	HepB_BD	Pol3	Hib3	MCV1
97.875	68.875	93.500	92.500	91.625

3. Descriptive Overview of California Vaccinations
    a. The mean levels of WithoutDTP, WithoutPolio, WithoutMMR, and WithoutHepB across districts





Mean rate of each of these four variables

WithoutDTP	WithoutPolio	WithoutMMR	WithoutHepB
10.21286	9.778571	10.17286	7.691429

b. Correlation Matrix

	WithoutDTP	WithoutPolio	WithoutMMR	WithoutHepB
WithoutDTP	1.00	0.98	0.98	0.89
WithoutPolio	0.98	1.00	0.97	0.91
WithoutMMR	0.98	0.97	1.00	0.90
WithoutHepB	0.89	0.91	0.90	1.00

According the table above, if students are missing one vaccine, it is very likely that they are missing all of the others.

c. Compared the WithoutDTP, WithoutPolio, WithoutMMR rate of U.S. Vaccination level, California rate was higher. Compared the WithoutHepB rate of U.S. Vaccination level, California rate was much lower.

4. In general, California's vaccination rate was lower than the level of U.S. California government should increase vaccine promotion for students, promote them to vaccine all kinds of vaccination.

Inferential Reporting

5.

a. Without variables transformation, PctFamilyPoverty predicts the percentage of all enrolled students with belief exceptions. The coefficient of PctFamilyPoverty is -0.334508 and P value is 2.82e-05
b. With variables transformation, PctFamilyPoverty, logEnrolled (Enrolled after transforming), and logTotSchools (TotalSchools after transforming) predict the percentage of all enrolled students with belief exceptions.
    i. The coefficient of logEnrolled is -2.86310 and P value is 3.6e-08
    ii. The coefficient of logTotSchools is 2.02452and P value is 0.00406
    iii. The P-value of the whole model is p-value: < 2.2e-16, Multiple R-squared is 0.1412

6.

a. Without variables transformation, PctFamilyPoverty, Enrolled, and TotalSchools predict the percentage of all enrolled students with up-to-date vaccines.

```
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       0.9835234  0.0187811  52.368  < 2e-16 ***
Enrolled          0.0002119  0.0000362   5.852 7.47e-09 ***
PctChildPoverty   0.0011812  0.0014092   0.838   0.4022
PctFamilyPoverty -0.0041439  0.0021076  -1.966   0.0497 *
TotalSchools     -0.0216205  0.0033464  -6.461 1.96e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2288 on 695 degrees of freedom
Multiple R-squared:  0.0987,    Adjusted R-squared:  0.09352
F-statistic: 19.03 on 4 and 695 DF,  p-value: 7.316e-15
```

b. With variables transformation, logEnrolled, PctFamilyPoverty and logTotSchools predict the percentage of all enrolled students with up-to-date vaccines. And R-Squared of this model is higher than the previous model.

```
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      63.99154    2.99700  21.352  < 2e-16 ***
logEnrolled       4.26336    0.73189   5.825 8.73e-09 ***
PctChildPoverty   0.09067    0.07414   1.223  0.22171
PctFamilyPoverty  0.25992    0.11074   2.347  0.01919 *
logTotSchools    -3.06518    1.00023  -3.064  0.00226 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.8 on 695 degrees of freedom
Multiple R-squared:  0.1477,    Adjusted R-squared:  0.1428
F-statistic:  30.1 on 4 and 695 DF,  p-value: < 2.2e-16
```

7. WithoutDTP, WithoutMMR, and WithoutHepB are the set of predictors that I can achieve the best R-squared in predicting the percentage of all enrolled students with completely . The R-squared is 94.41%.
8. There is an interaction between PctChildPoverty and Enrolled

```
Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              8.022e+01  1.167e+00  68.715  < 2e-16 ***
PctChildPoverty          3.008e-01  4.366e-02   6.890 1.25e-11 ***
Enrolled                 5.711e-03  1.326e-03   4.307 1.90e-05 ***
PctChildPoverty:Enrolled -1.797e-04  4.362e-05  -4.120 4.25e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.33 on 696 degrees of freedom
Multiple R-squared:  0.06915,   Adjusted R-squared:  0.06514
```

9. Enrolled and TotalSchools predict whether or not a district’s reporting was complete.

```
Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       3.3706292  0.3825464   8.811  < 2e-16 ***
PctChildPoverty   0.0230958  0.0284693   0.811 0.417221
PctFamilyPoverty -0.0673465  0.0391132  -1.722 0.085099 .
Enrolled          0.0019645  0.0007104   2.765 0.005689 **
TotalSchools     -0.1972429  0.0591322  -3.336 0.000851 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

10. For improving vaccination rate and reporting compliance, California government should pay more attention to the percentage of students with completely up-to-date vaccines and decrease the percentage of families in district living below the poverty line.