

Code for Full Statistical Analysis

2024-03-28

Data Cleaning

Check School Types

```
## 'summarise()' has grouped output by 'year'. You can override using the  
## '.groups' argument.
```

```
## # A tibble: 7 x 3  
##   type      '2017-18' '2018-19'  
##   <chr>      <int>      <int>  
## 1 BOCES          47         NA  
## 2 Kindergarten  1302         NA  
## 3 Nonpublic     166         NA  
## 4 Private      1826        3084  
## 5 Public       2976        9563  
## 6 Unknown      7880       19294  
## 7 Charter        NA         273
```

Regroup school type by assigning BOCES, kindergarten, nonpublic to “Other” category.

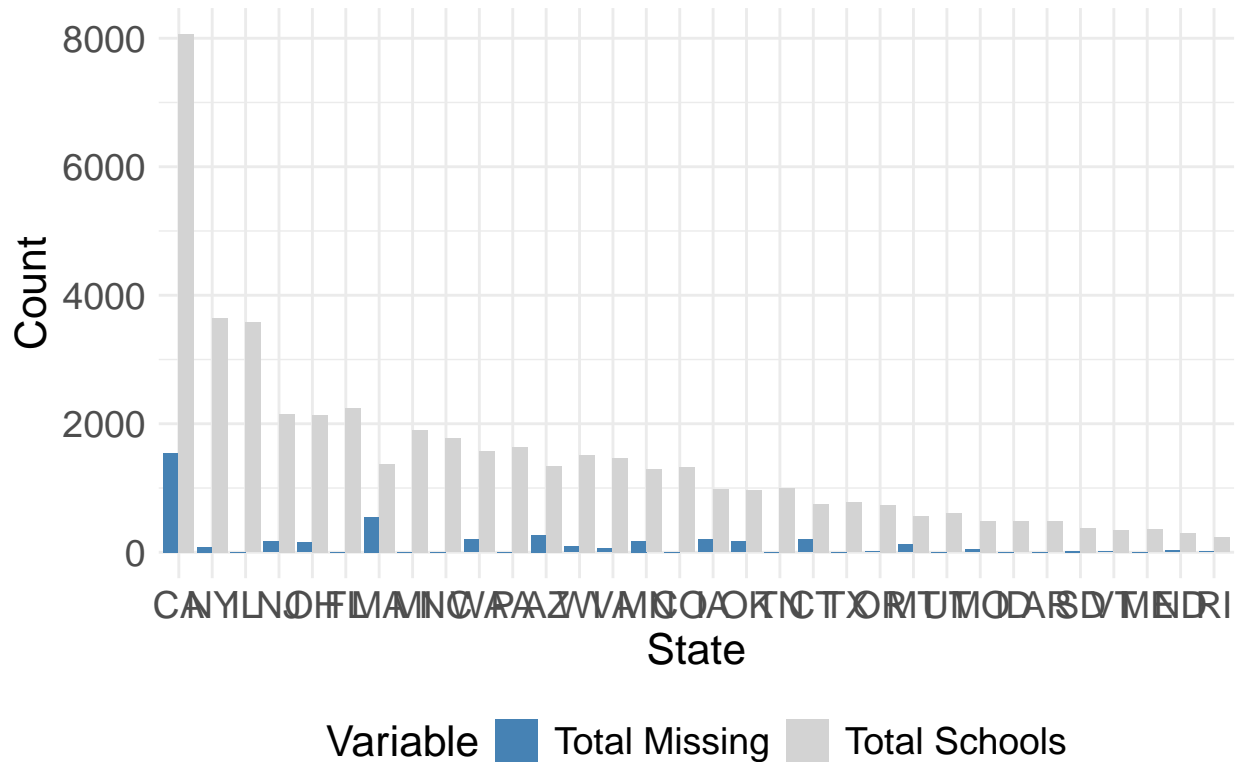
Check missings in vaccination rate

```
## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be  
## resized.
```

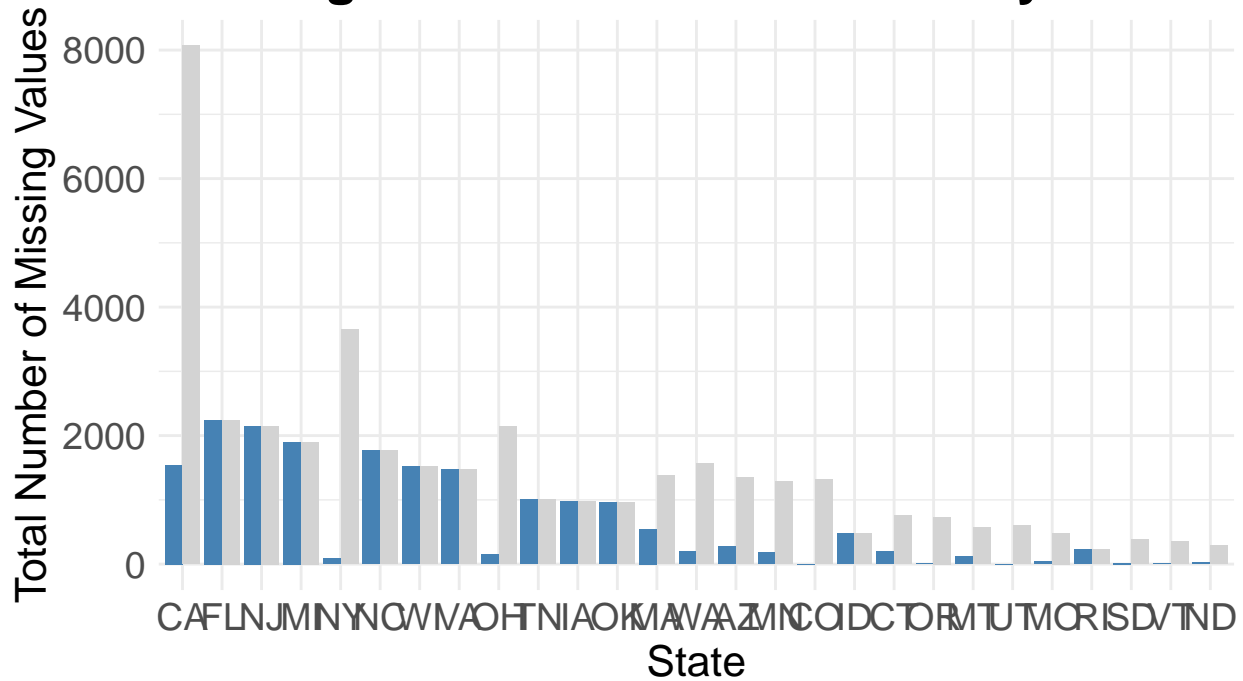
State	Number of Missing in Vaccination Rate			Total No.Schools
	MMR	Overall	Total	
Colorado	2	4	6	1319
Oregon	10	10	20	733
Vermont	11	11	22	349
Rhode Island	230	15	245	230
Ohio	149	151	300	2137
North Dakota	27	293	320	293
Maine	0	353	353	353
South Dakota	17	380	397	380
Washington	199	199	398	1567
Arkansas	0	481	481	481
Idaho	475	8	483	475
Missouri	43	484	527	484

Utah	1	602	603	602
Montana	119	567	686	567
Texas	0	777	777	777
Connecticut	197	751	948	751
Tennessee	1004	0	1004	1004
Oklahoma	965	171	1136	965
Iowa	978	195	1173	978
Minnesota	177	1289	1466	1289
Virginia	1468	55	1523	1468
Wisconsin	1515	92	1607	1515
Arizona	270	1343	1613	1343
Pennsylvania	0	1635	1635	1635
North Carolina	1767	1	1768	1767
Michigan	1899	0	1899	1899
Massachusetts	547	1374	1921	1374
Florida	2242	6	2248	2242
New Jersey	2149	167	2316	2149
California	1543	1544	3087	8066
Illinois	0	3574	3574	3574
New York	83	3645	3728	3645

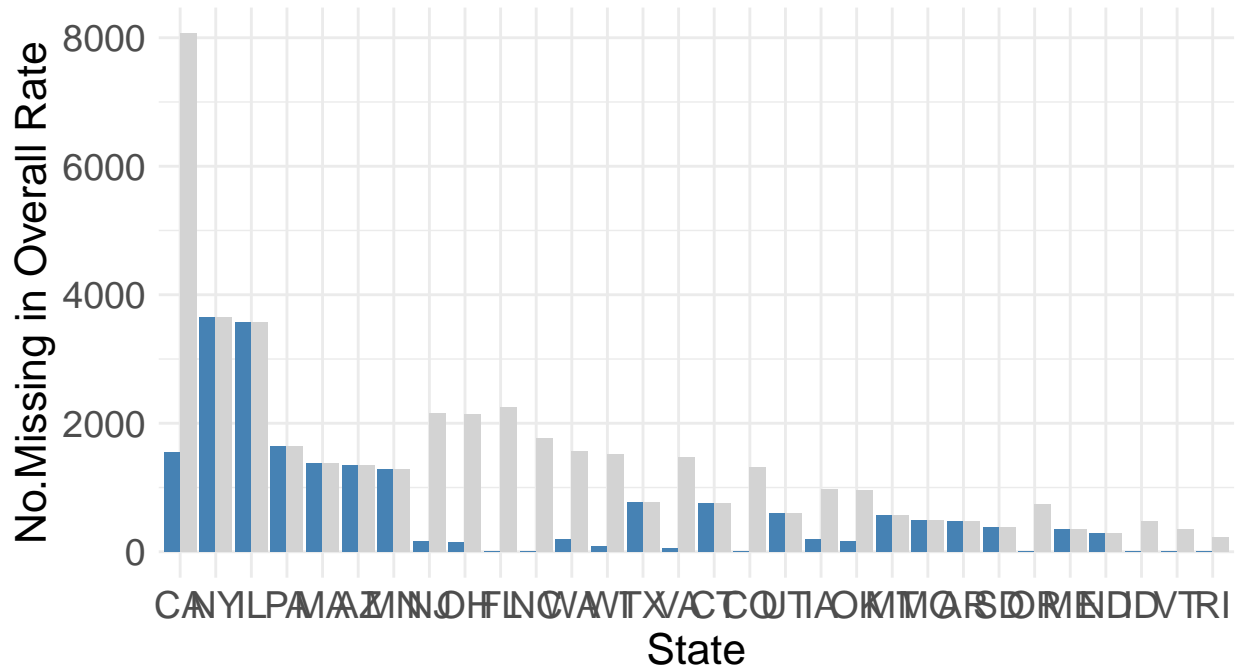
Comparison of Total Missing Data and Total Schools



Missing MMR Vaccination Rates by State



Missing Overall Vaccination Rates by State



Variable ■ Missing in Overall ■ Total Schools

Check Duplicates

```
## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be  
## resized.
```

Table 2: Sample Duplicated School Observations in California

School ID	No.Record	School Name	County	City	Type	Year	Enrollment	MMR Rate	Overall Rate
5079	1	Stratford	Alameda	Fremont	Private	2018-19	88	98	98
7313	2	Stratford	Alameda	Fremont	Private	2018-19	45	95	95
2454	1	Lincoln Elementary	Fresno	Fresno	Public	2018-19	108	99	98
4232	2	Lincoln Elementary	Fresno	Fresno	Public	2018-19	74	98	98
8181	1	Anneliese Schools	Orange	Laguna Beach	Private	2018-19	28	79	64
8198	2	Anneliese Schools	Orange	Laguna Beach	Private	2018-19	39	77	77
5078	1	Stratford	Santa Clara	Santa Clara	Private	2018-19	161	98	98
7316	2	Stratford	Santa Clara	Santa Clara	Private	2018-19	43	95	95
5085	1	Stratford	Santa Clara	Sunnyvale	Private	2018-19	83	98	98
5086	2	Stratford	Santa Clara	Sunnyvale	Private	2018-19	88	98	98

Table 1

```
## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
## resized.
```

State	N (Schools)	N (City)	N (County)	School-Aged Population (2016)	Expenditure (2016)	Per Cap
Arizona	1343	178	15	1199	7137123	5952.5
Arkansas	481	246	1	515	4411761	8566.5
California	8061	987	58	6602	68892072	10435.0
Colorado	1313	1	66	928	8060420	8685.7
Connecticut	751	169	8	568	8687640	15295.1
Florida	2092	1	1	3037	24222830	7975.9
Idaho	475	175	46	323	1935653	5992.7
Illinois	3569	808	104	2152	23933484	11121.5
Iowa	977	1	99	533	5517359	10351.5
Maine	353	1	16	190	2290535	12055.4
Massachusetts	1374	377	14	1016	13998163	13777.7
Michigan	1889	1	83	1621	15328155	9455.9
Minnesota	1287	1	88	940	9987116	10624.5
Missouri	484	194	1	1012	8996869	8890.1
Montana	567	1	56	165	1598517	9687.9
New Jersey	2135	1	22	1451	24349836	16781.4
New York	3638	965	62	2982	56656536	18999.5
North Carolina	1767	1	100	1690	12620939	7468.0
North Dakota	293	168	53	121	1393844	11519.3
Ohio	2137	634	88	1920	20271944	10558.3
Oklahoma	965	408	78	697	4958496	7114.0
Oregon	733	1	36	637	6285584	9867.4
Pennsylvania	1635	669	1	1965	25955384	13208.8
Rhode Island	230	37	1	154	2297516	14918.9
South Dakota	380	1	65	152	1185796	7801.2
Tennessee	1003	1	95	1098	8294165	7553.8
Texas	777	1	93	5310	43287224	8152.0
Utah	602	164	29	668	3931265	5885.1
Vermont	349	1	15	88	1536843	17464.1
Virginia	1468	1	130	1358	14385736	10593.3
Washington	1567	329	38	1181	12315163	10427.7
Wisconsin	1515	470	73	952	9691073	10179.6

```
## # A tibble: 1 x 5
##   N_both Ncityonly Ncountyonly N_none Nothers
##   <dbl>   <dbl>       <dbl> <dbl>   <dbl>
## 1     11       4         13     1     13
```

Filter dataframes for MMR and Overall Vaccination Rate

Stratified Sampling

MMR

Cleaning for Overall Vaccination

EDA

State School number

```
## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
## resized.
```

Table 4: The number of schools within each state.

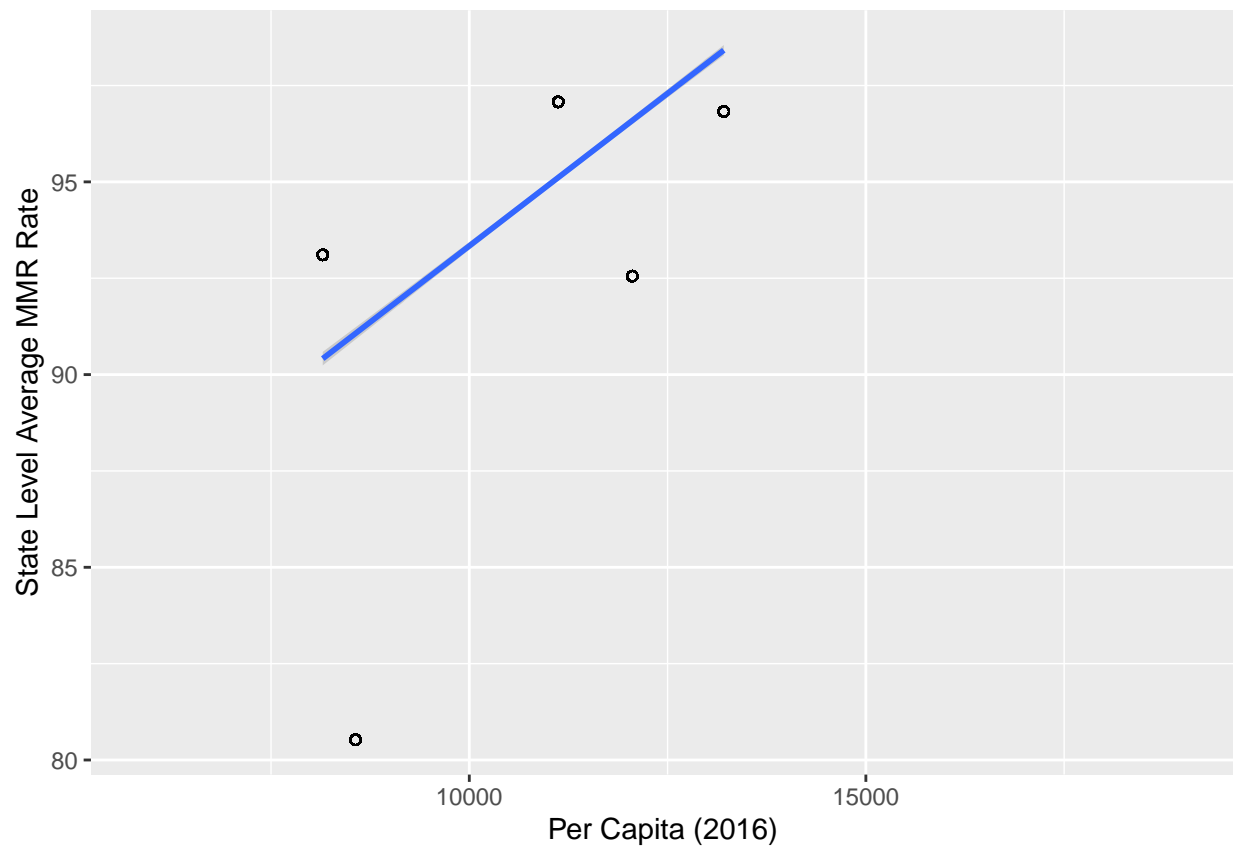
State	Schools (N)
Rhode Island	230
North Dakota	270
Vermont	349
Maine	351
South Dakota	376
Missouri	430
Arkansas	458
Idaho	475
Montana	537
Utah	601
Oregon	698
Connecticut	732
Texas	763
Iowa	878
Oklahoma	881
Tennessee	954
Minnesota	1176
Colorado	1246
Arizona	1301
Massachusetts	1301
Wisconsin	1302
Washington	1398
Virginia	1435
Pennsylvania	1536
North Carolina	1664
Michigan	1770
Ohio	1913
Florida	2092
New Jersey	2131
Illinois	3112
New York	3465
California	6891

Data Visualization

```
## 'geom_smooth()' using formula = 'y ~ x'
```

```
## Warning: Removed 24879 rows containing non-finite outside the scale range
## ('stat_smooth()').
```

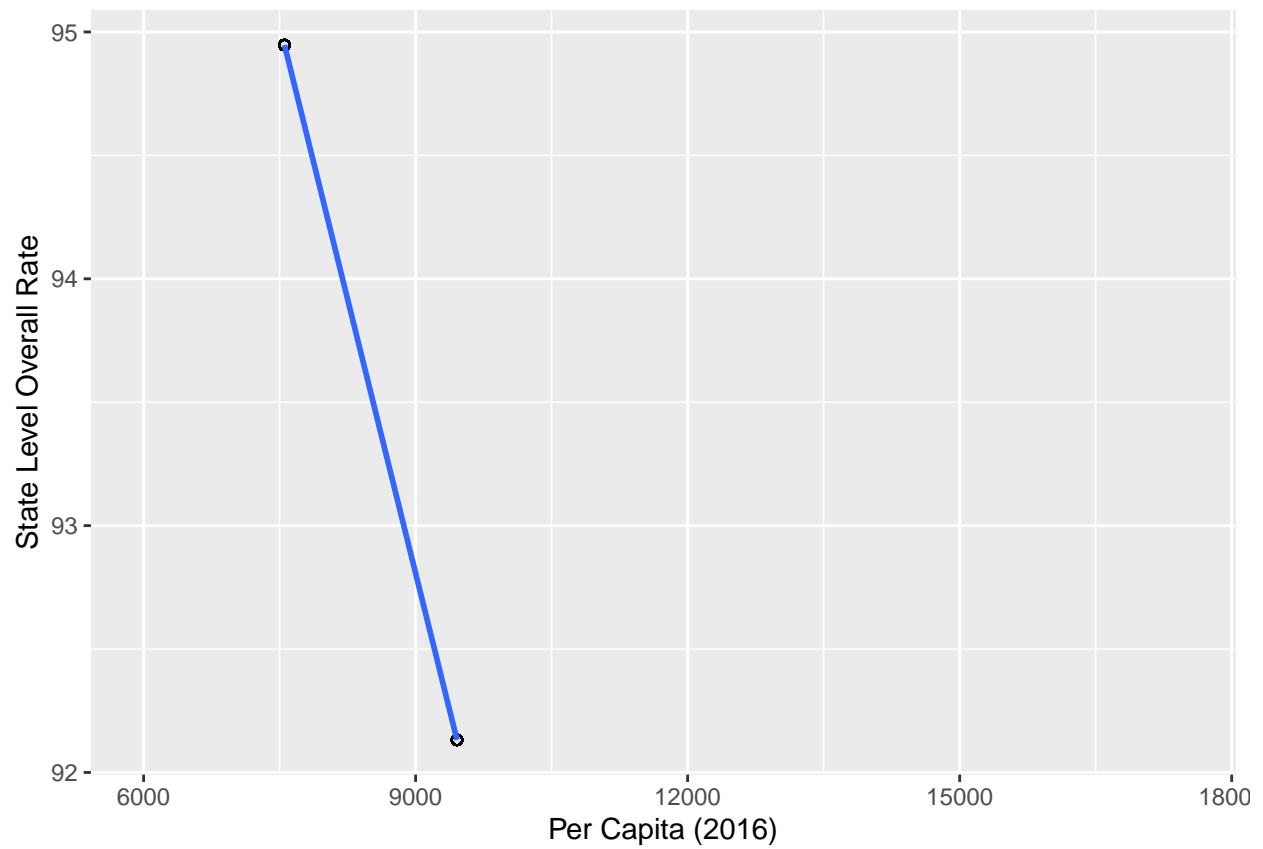
```
## Warning: Removed 24879 rows containing missing values or values outside the scale range
## ('geom_point()').
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```

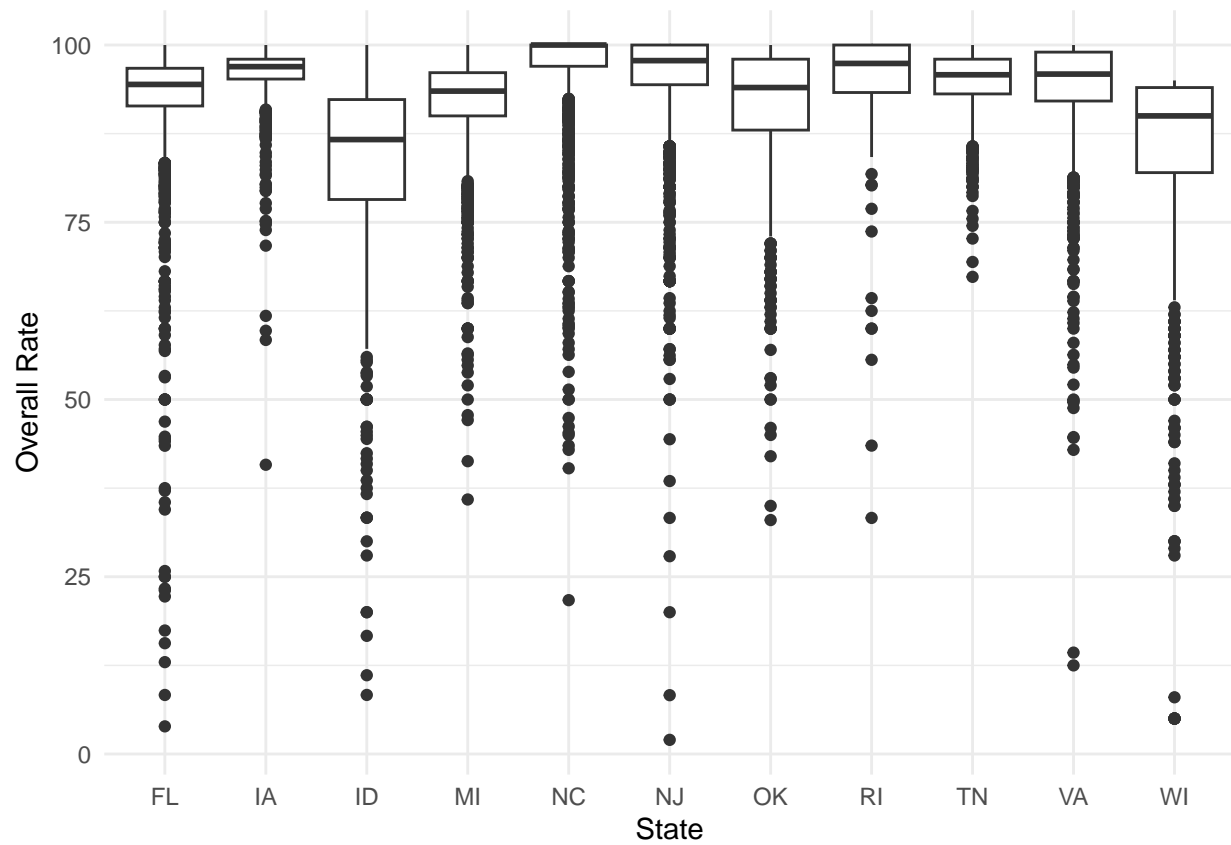
```
## Warning: Removed 25784 rows containing non-finite outside the scale range  
## ('stat_smooth()').
```

```
## Warning: Removed 25784 rows containing missing values or values outside the scale range  
## ('geom_point()').
```

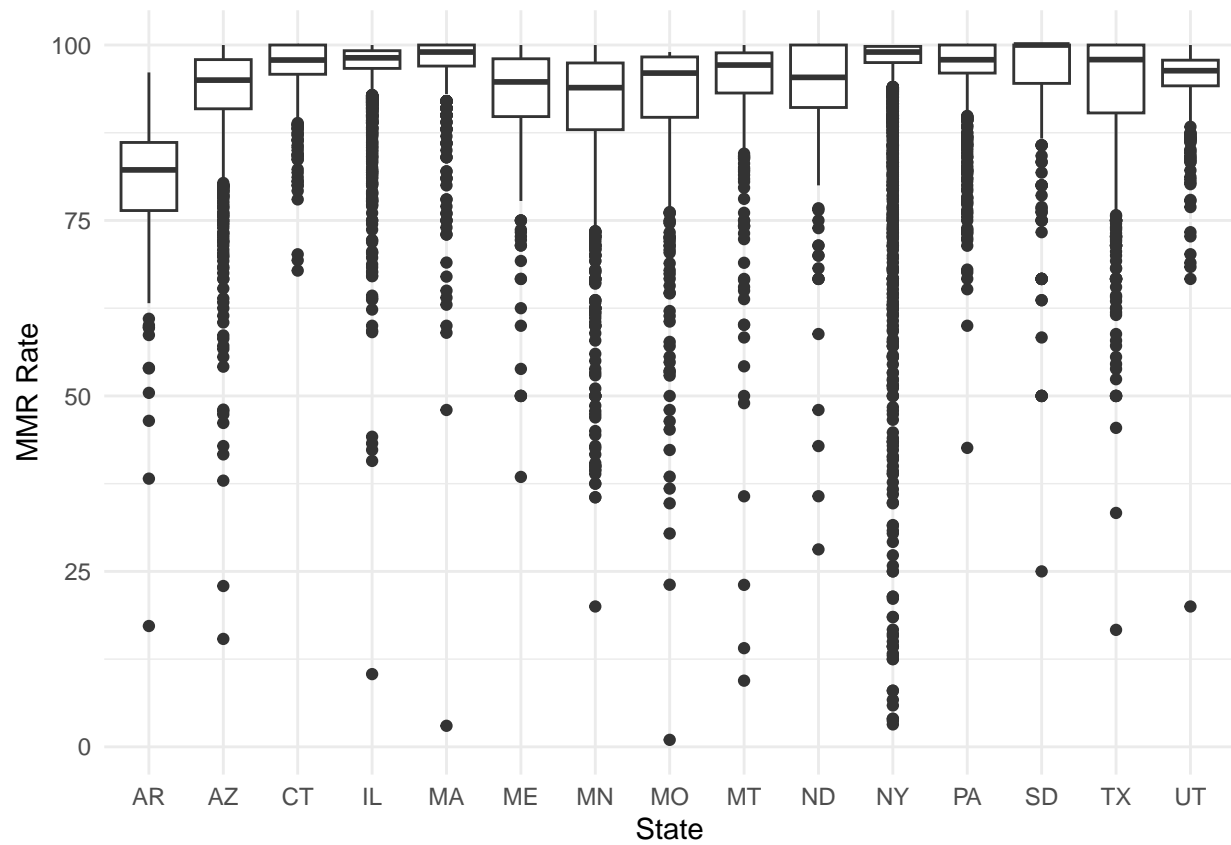


Vaccination Rates by States

```
## Warning: Removed 15226 rows containing non-finite outside the scale range  
## ('stat_boxplot()').
```

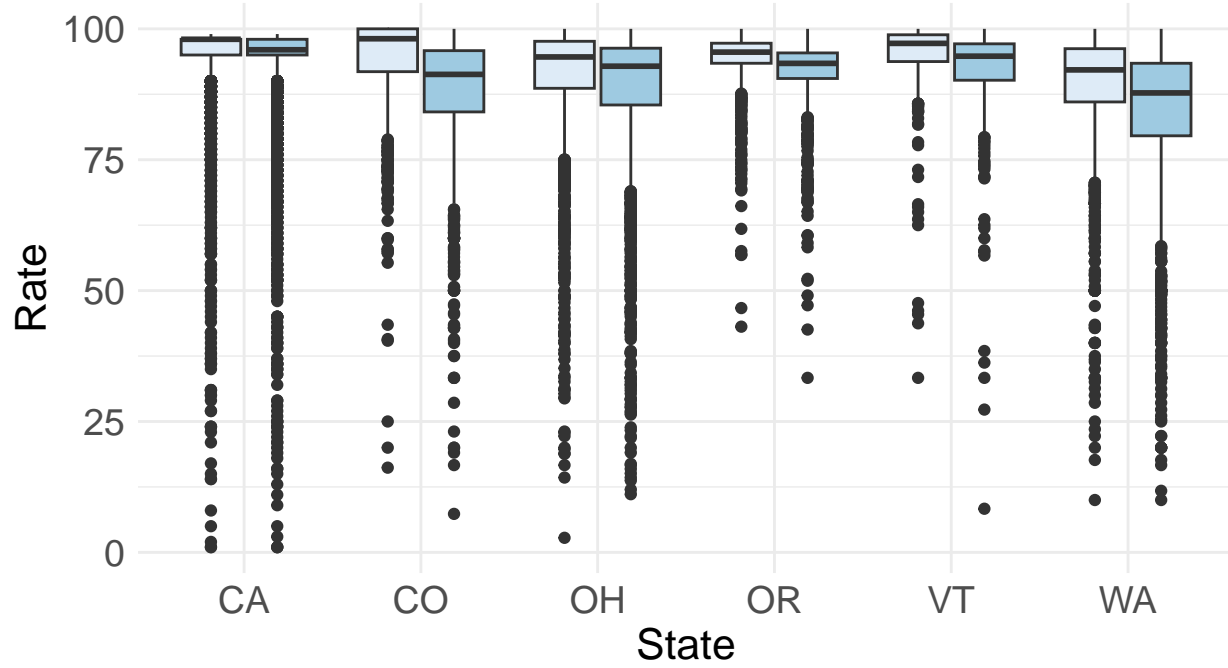


```
## Warning: Removed 19015 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```



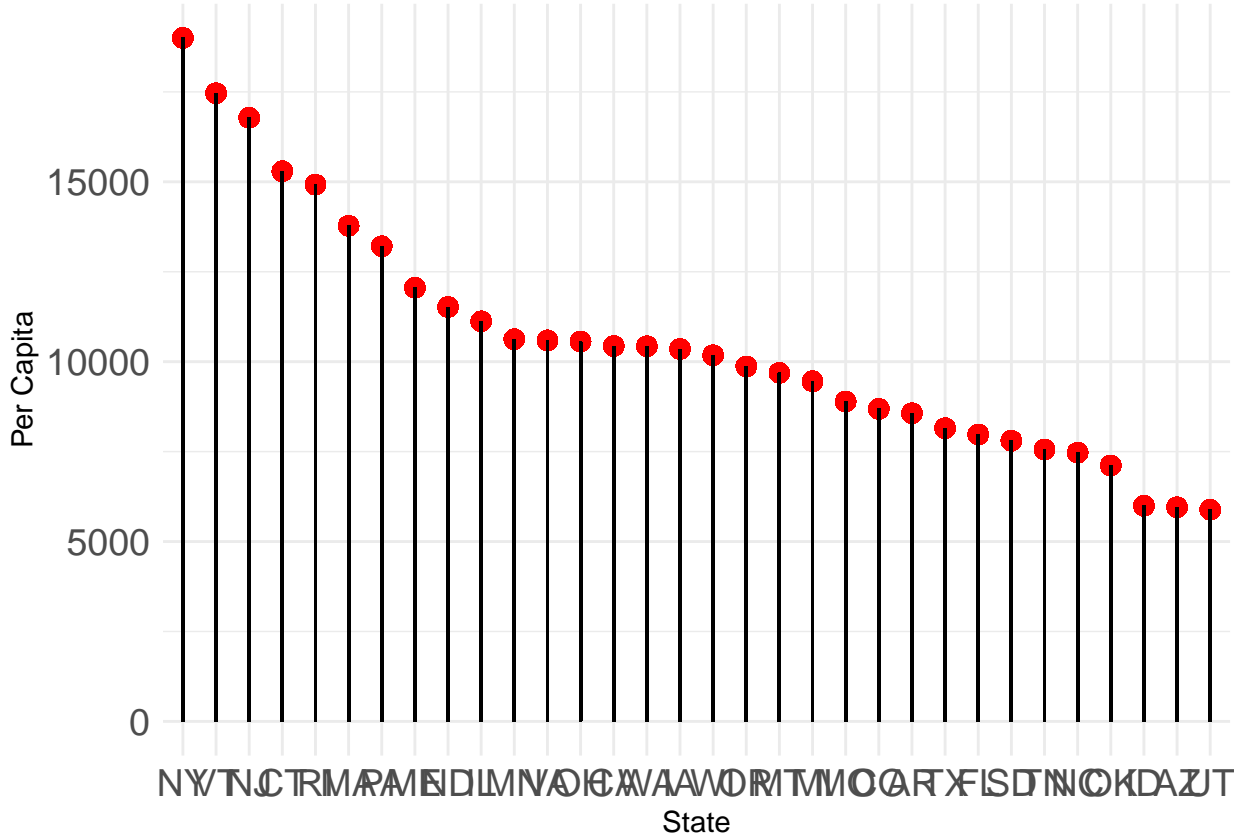
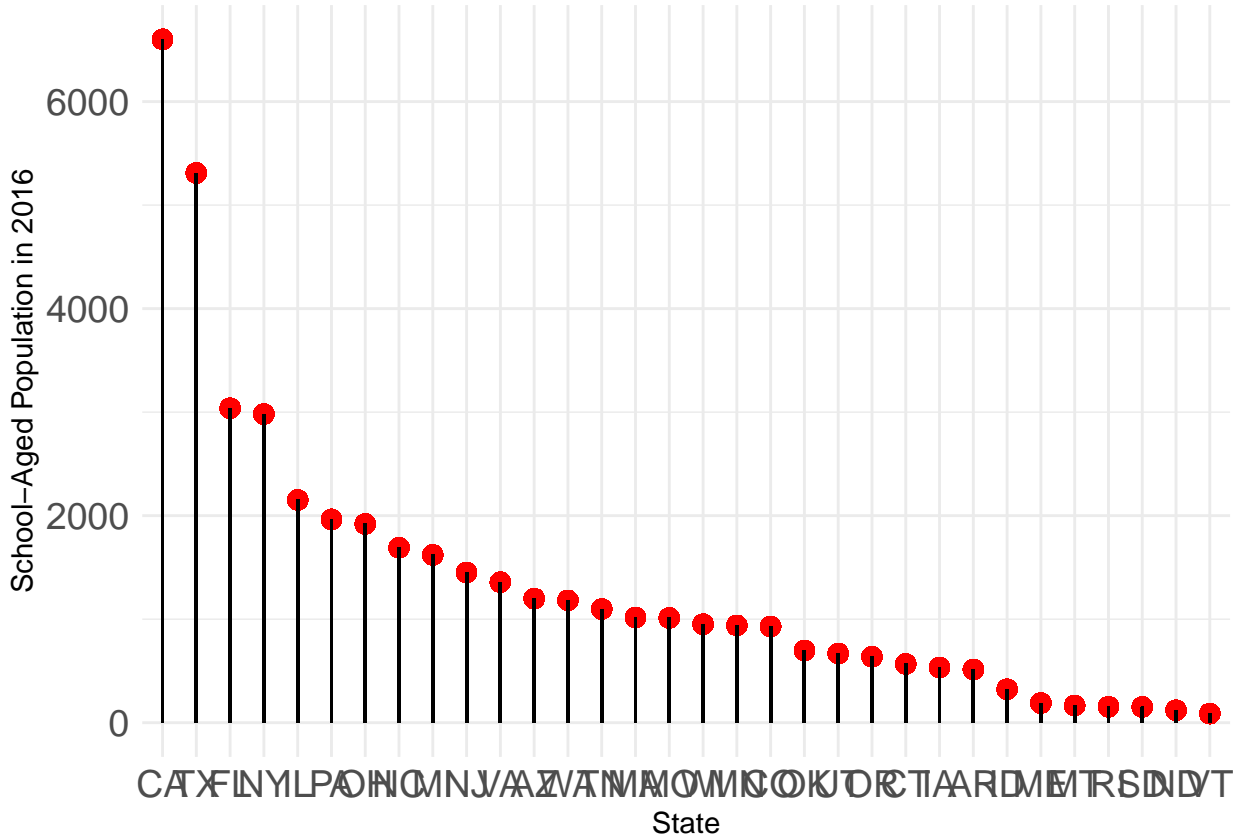
```
## Warning: Removed 3833 rows containing non-finite outside the scale range
## ('stat_boxplot()').
```

Comparison of Vaccination Rate Distribution



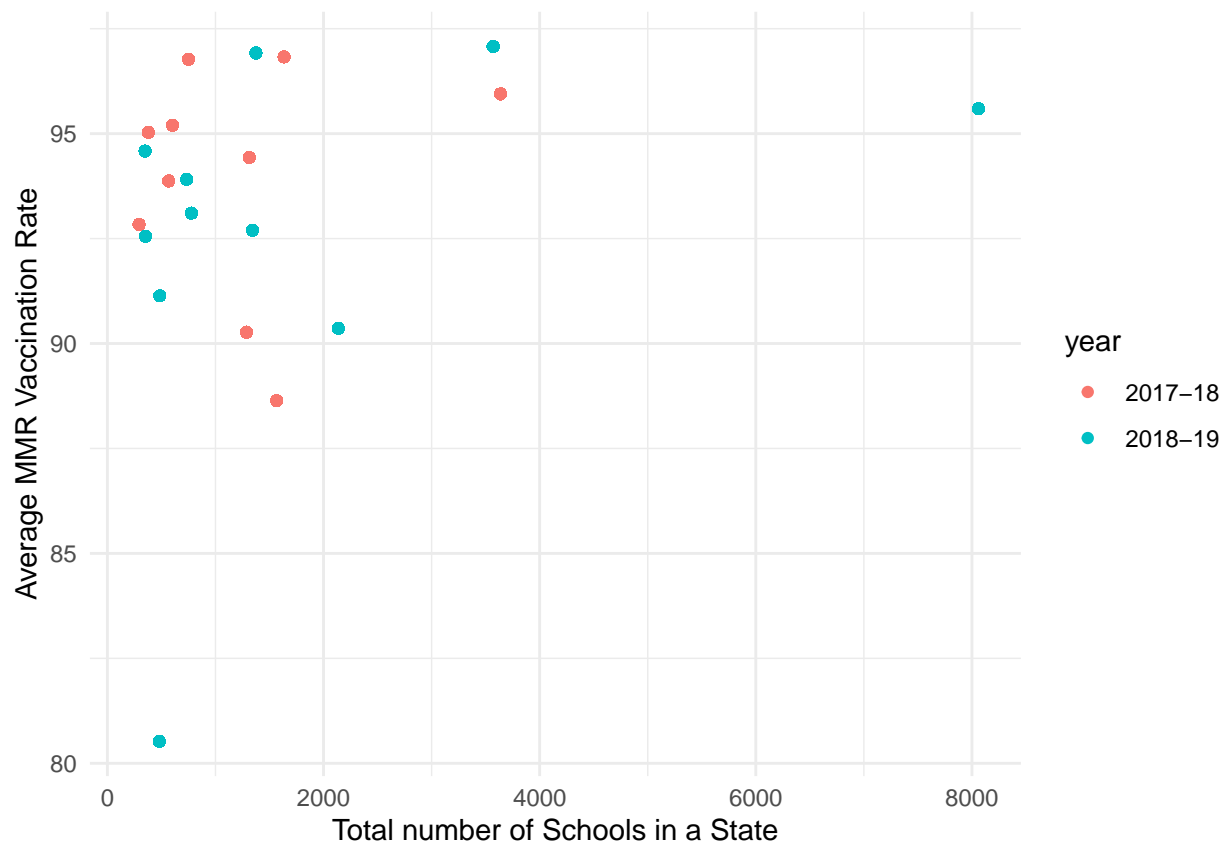
Vaccine  MMR  Overall

State School Aged Population and Per Capita

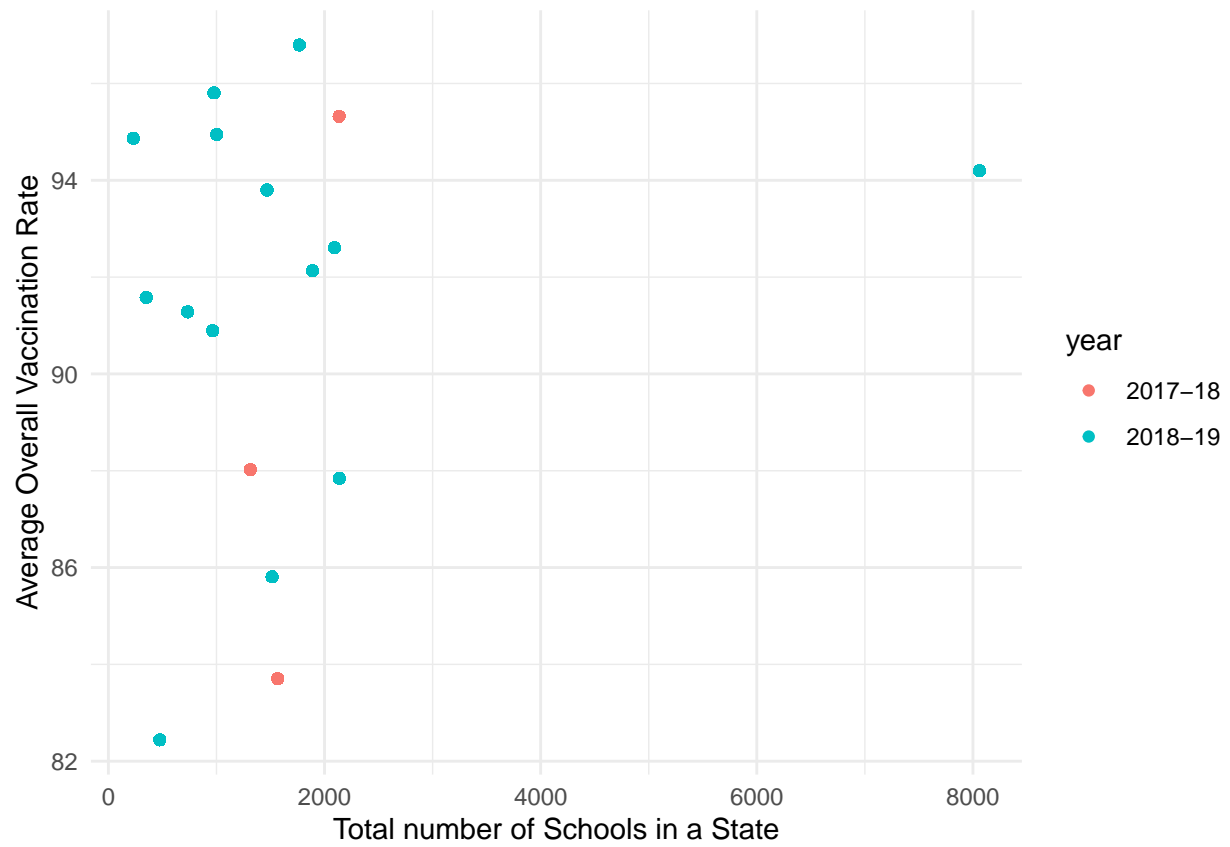


Vaccination rate vs. School, County, City, Capita, and Spending

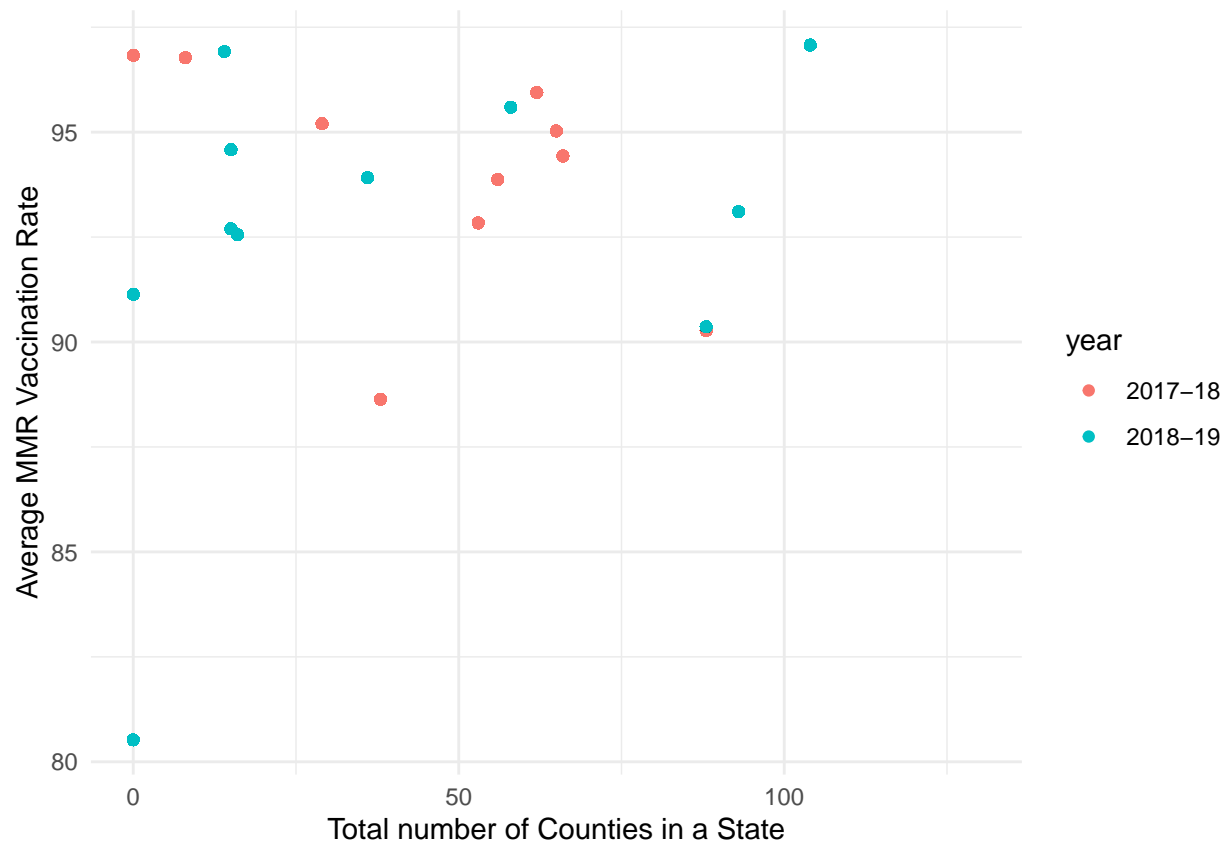
```
## Warning: Removed 14516 rows containing missing values or values outside the scale range  
## ('geom_point()').
```



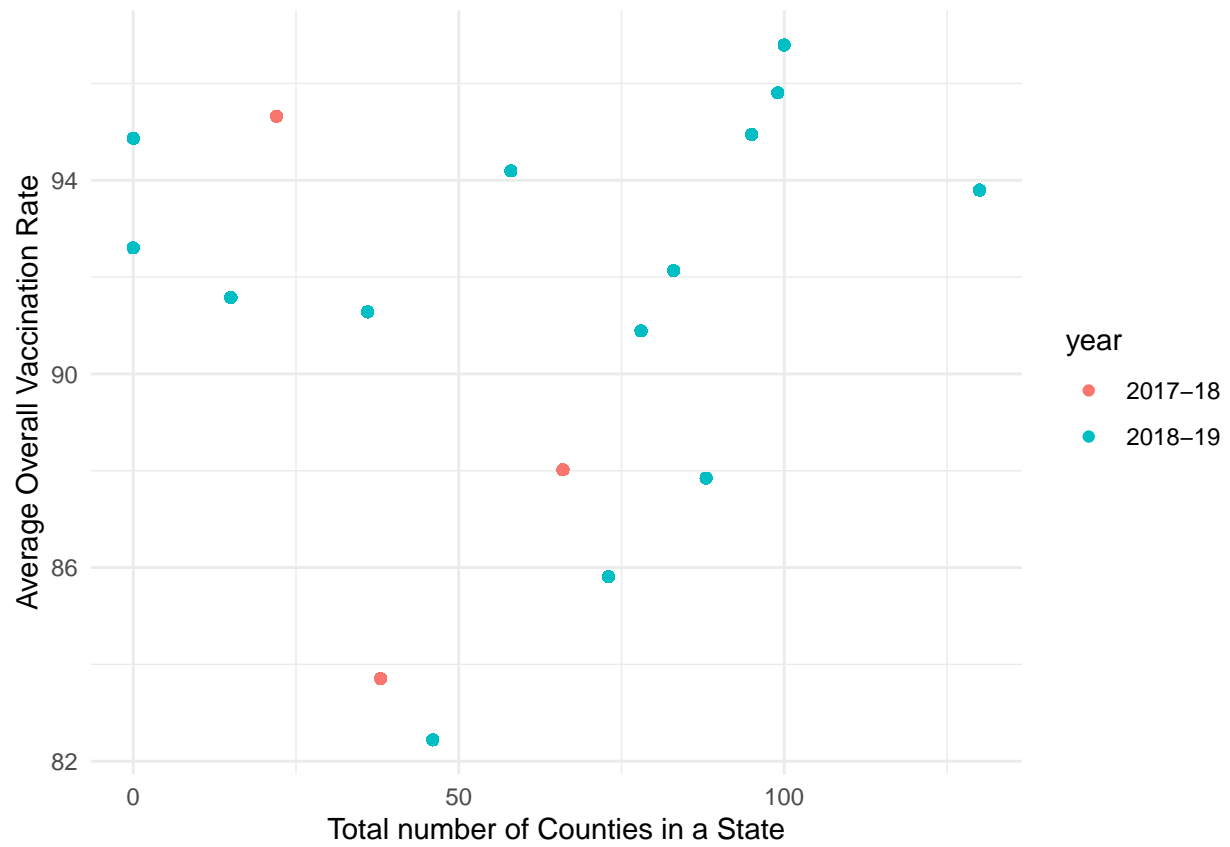
```
## Warning: Removed 17534 rows containing missing values or values outside the scale range  
## ('geom_point()').
```



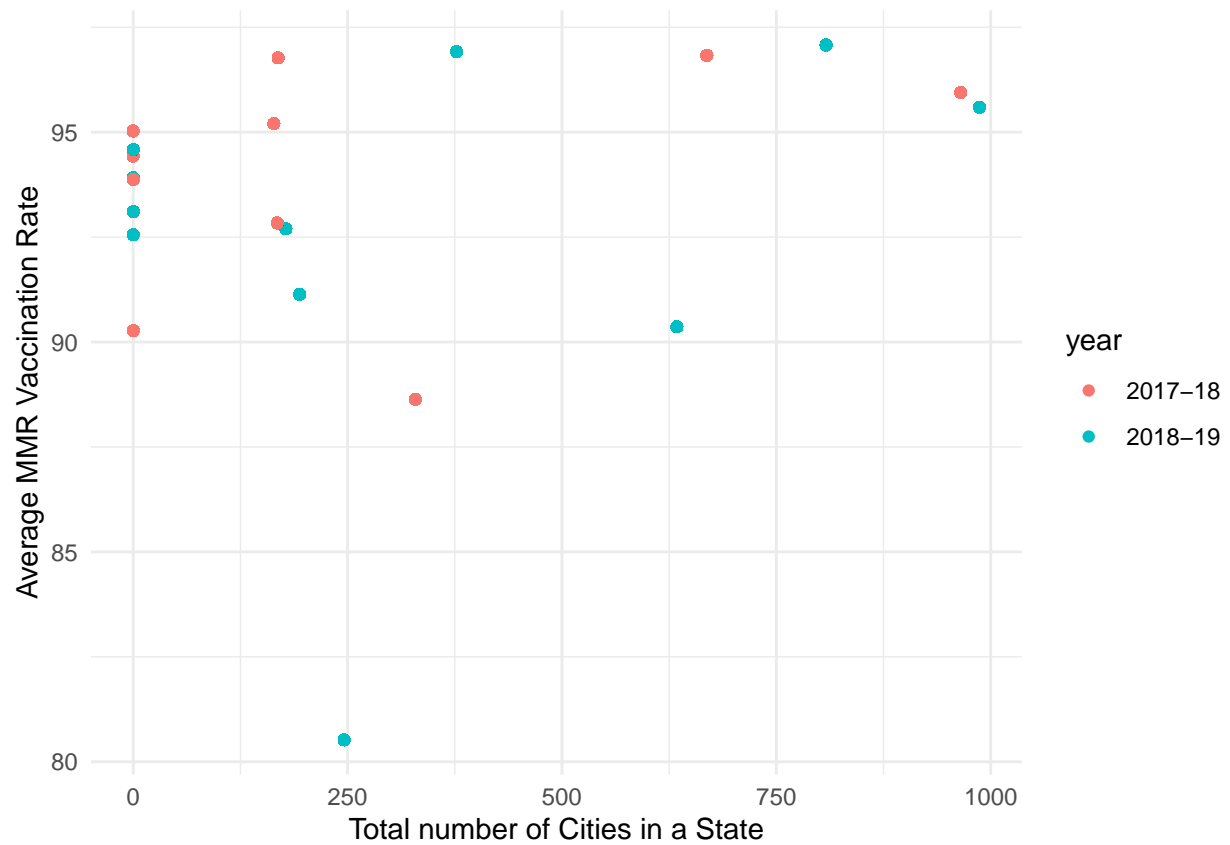
```
## Warning: Removed 14516 rows containing missing values or values outside the scale range
## ('geom_point()').
```

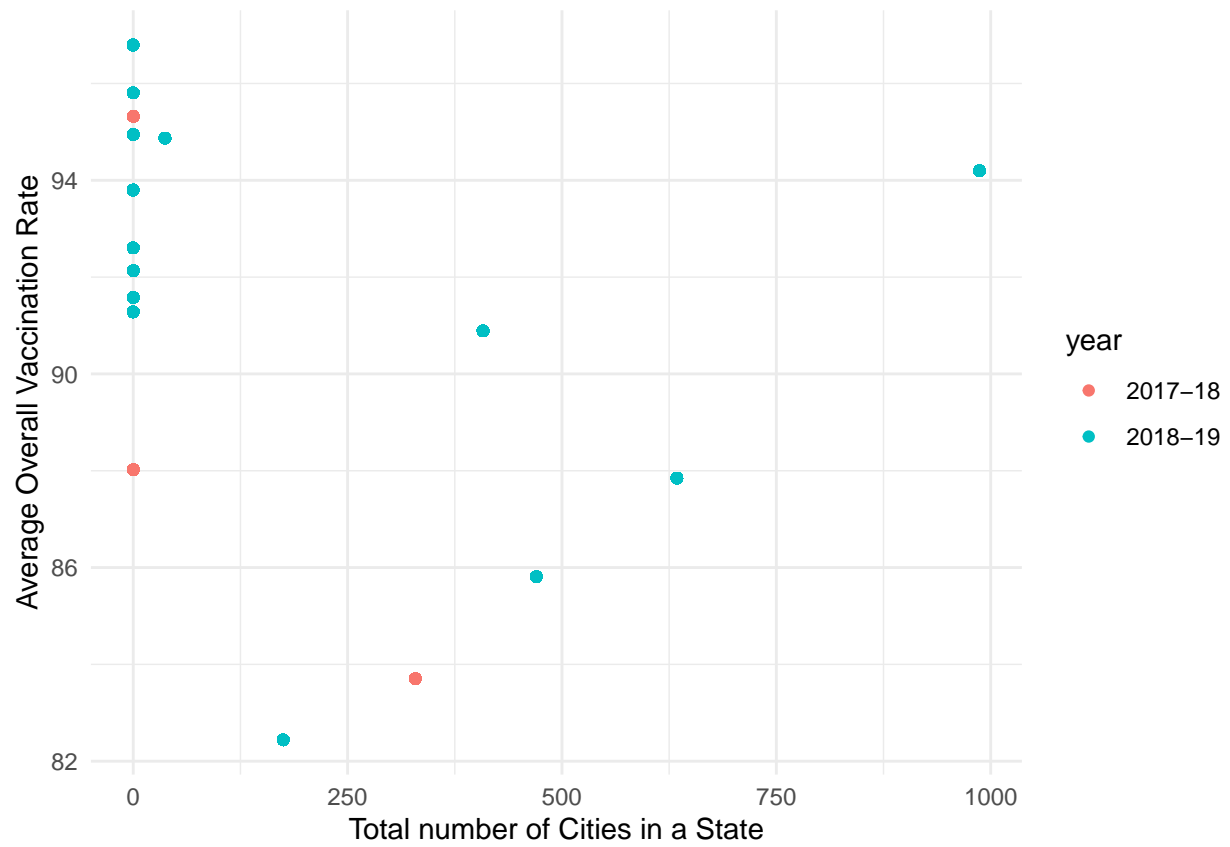
```
## Warning: Removed 17534 rows containing missing values or values outside the scale range
## ('geom_point()').
```



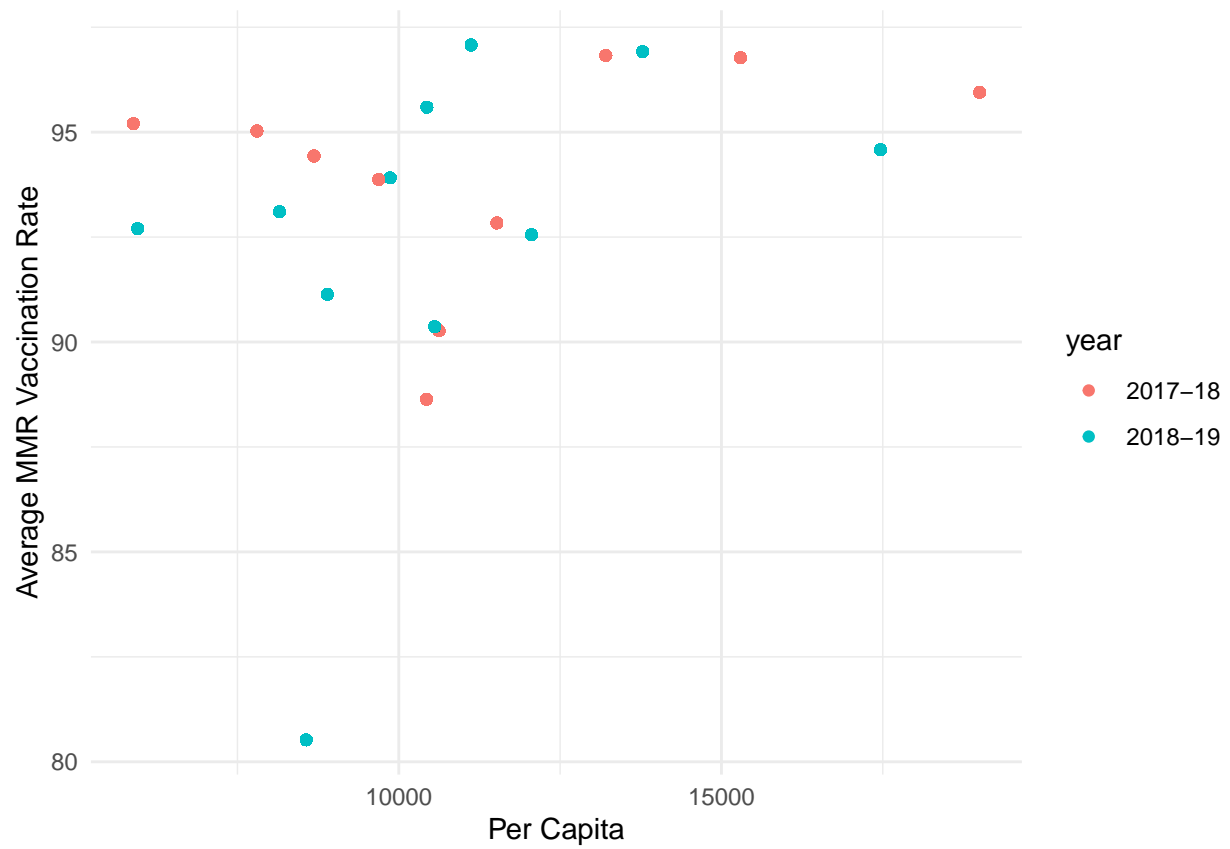
```
## Warning: Removed 14516 rows containing missing values or values outside the scale range
## ('geom_point()').
```



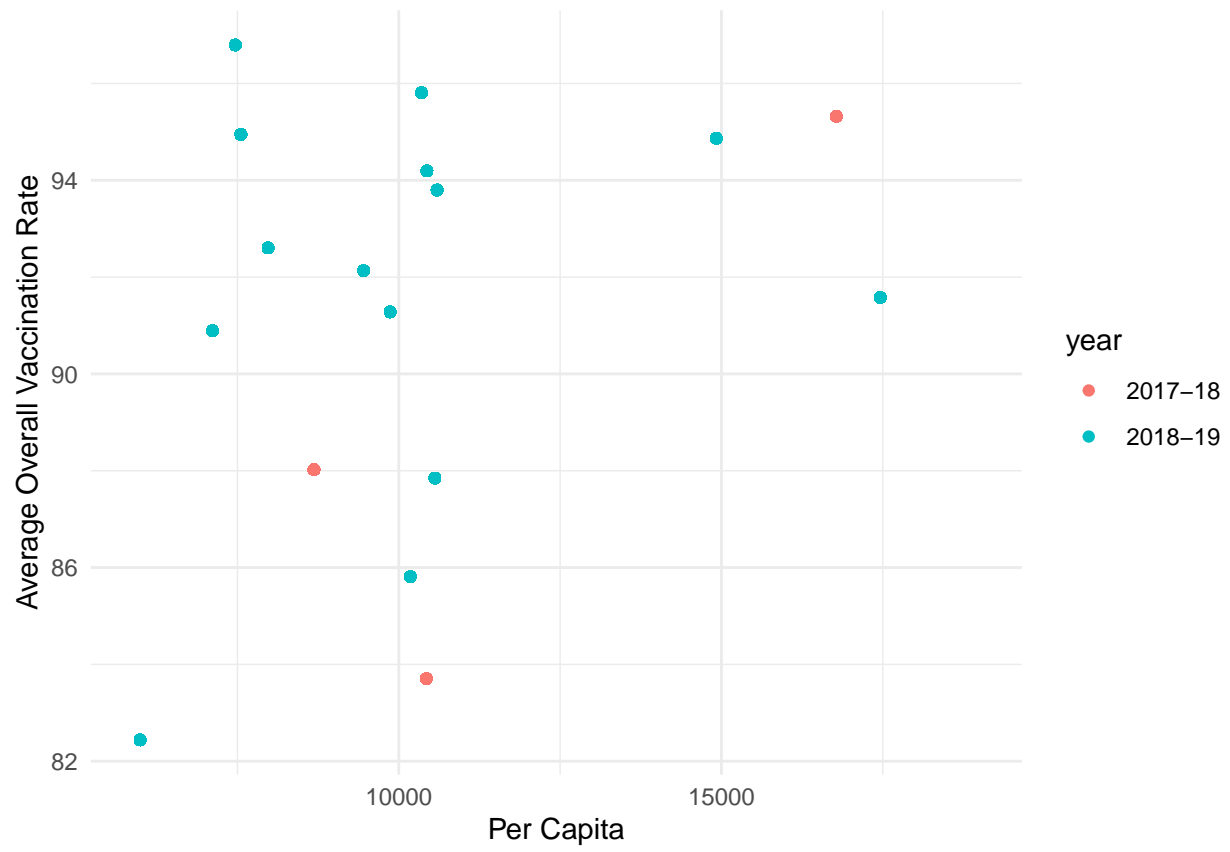
```
## Warning: Removed 17534 rows containing missing values or values outside the scale range
## ('geom_point()').
```



```
## Warning: Removed 14516 rows containing missing values or values outside the scale range
## ('geom_point()').
```



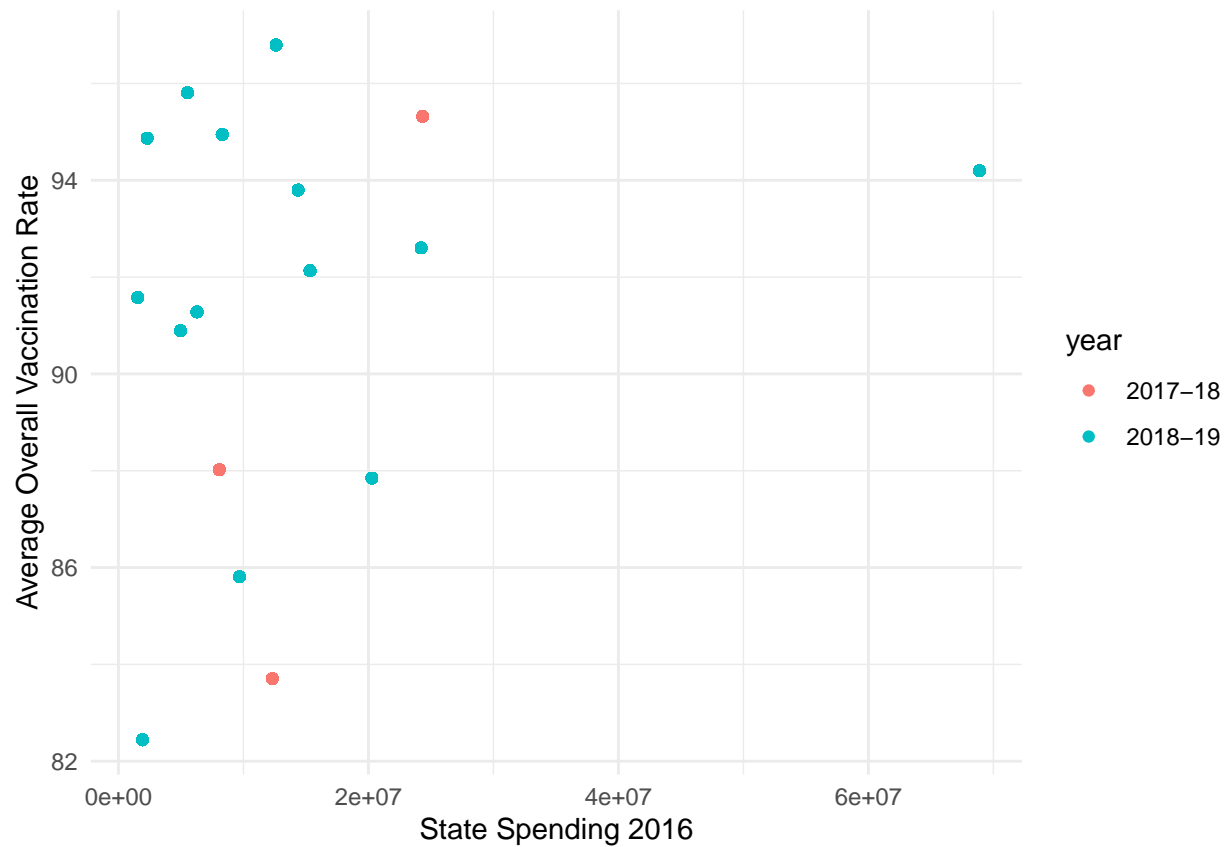
```
## Warning: Removed 17534 rows containing missing values or values outside the scale range
## ('geom_point()').
```



```
## Warning: Removed 14516 rows containing missing values or values outside the scale range
## ('geom_point()').
```



```
## Warning: Removed 17534 rows containing missing values or values outside the scale range
## ('geom_point()').
```



Florida

```
## overall
## Min. : 3.90
## 1st Qu.: 91.40
## Median : 94.44
## Mean : 92.60
## 3rd Qu.: 96.72
## Max. : 100.00
## NA's : 6
```

EDA for MMR Data

```
## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
## resized.
```

Table 5: The summary of MMR vaccination rates

State	No.School	Min	Mean	Median	Max	No.Missing	School Year
Colorado	1246	16.18000	94.43079	98.11000	100.00000	2	2017-18
Connecticut	732	67.86000	96.77105	97.86000	100.00000	197	2017-18
Minnesota	1176	20.00000	90.27094	93.92500	100.00000	177	2017-18
Montana	537	9.43000	93.87281	97.14500	100.00000	119	2017-18
New York	3465	3.20000	95.94616	99.00000	100.00000	83	2017-18

North Dakota	270	28.13000	92.83635	95.39000	100.00000	27	2017-18
Pennsylvania	1536	42.60000	96.82862	97.90000	100.00000	0	2017-18
South Dakota	376	25.00000	95.02821	100.00000	100.00000	17	2017-18
Utah	601	20.00000	95.20240	96.34000	100.00000	1	2017-18
Washington	1398	10.00000	88.63814	92.16000	100.00000	199	2017-18
Arizona	1301	15.38000	92.69992	95.00000	100.00000	270	2018-19
Arkansas	458	17.21854	80.52271	82.20472	96.09484	0	2018-19
California	6891	1.00000	95.59251	98.00000	99.00000	1543	2018-19
Illinois	3112	10.36585	97.07555	98.17579	100.00000	0	2018-19
Maine	351	38.46000	92.55482	94.74000	100.00000	0	2018-19
Massachusetts	1301	3.00000	96.92019	99.00000	100.00000	547	2018-19
Missouri	430	1.00000	91.13583	96.00000	99.00000	43	2018-19
Ohio	1913	2.78000	90.36530	94.63000	100.00000	149	2018-19
Oregon	698	43.11111	93.91200	95.54974	100.00000	10	2018-19
Texas	763	16.67000	93.10616	97.92000	100.00000	0	2018-19
Vermont	349	33.33333	94.58332	97.22222	100.00000	11	2018-19

EDA for Overall Data

```
## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
## resized.
```

Table 6: The summary of Overall vaccination rates

State	No.School	Min	Mean	Median	Max	No.Missing	School Year
Colorado	1246	7.350000	88.02157	91.30000	100	4	2017-18
New Jersey	2131	2.000000	95.32053	97.80000	100	167	2017-18
Washington	1398	10.000000	83.70287	87.76000	100	199	2017-18
California	6891	1.000000	94.19534	96.00000	99	1544	2018-19
Florida	2092	3.900000	92.60429	94.44000	100	6	2018-19
Idaho	475	8.333333	82.44082	86.66667	100	8	2018-19
Iowa	878	40.800000	95.80691	96.95000	100	195	2018-19
Michigan	1770	35.900000	92.13213	93.50000	100	0	2018-19
North Carolina	1664	21.700000	96.79428	100.00000	100	1	2018-19
Ohio	1913	11.110000	87.84588	92.86000	100	151	2018-19
Oklahoma	881	33.000000	90.89043	94.00000	100	171	2018-19
Oregon	698	33.333333	91.28114	93.40102	100	10	2018-19
Rhode Island	230	33.300000	94.86791	97.40000	100	15	2018-19
Tennessee	954	67.300000	94.94706	95.80000	100	0	2018-19
Vermont	349	8.333333	91.57887	94.77955	100	11	2018-19
Virginia	1435	12.500000	93.79837	95.90000	100	55	2018-19
Wisconsin	1302	5.000000	85.81096	90.00000	95	92	2018-19

Statistical Analysis

The data has many schools within each state, meaning that each state has repeated measurements by schools. Therefore, the vaccination rate between schools within the same state are not independent of each

other, violating the assumption of independence of standard regression models. Therefore, we will model the covariance structures with Generalized Least Squares.

MMR

Let Y_{ij} denote the MMR vaccination rate of the i^{th} state and j^{th} school.

$$E(Y_{ij}) = \beta_0 + \beta_1 \text{Capita}_i + \beta_2 \text{NSchool}_i + \beta_3 \text{Year}_i$$

where Capita_i was calculated by dividing the state spending by school aged population, NSchool represents the number of schools in each state, and Year denotes the school-year that the observation is recorded (1 = 2018-19, 0 = 2017-18).

Generalized Least Squares

Failed

Unadjusted GEE

$$\begin{aligned} \text{MMR}_{ij} &= \beta_0 + \beta_1 \cdot \text{PCS}_{ij} + \beta_2 \cdot \text{NSch}_{ij} + \beta_3 \cdot \text{Type}_{ij} + \epsilon_{ij} \\ \text{OVR}_{ij} &= \beta_0 + \beta_1 \cdot \text{PCS}_{ij} + \beta_2 \cdot \text{NSch}_{ij} + \beta_3 \cdot \text{Type}_{ij} + \epsilon_{ij} \end{aligned}$$

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

```
## Models are identical
```

```
## NULL
```

```
## Models are identical
```

```
## NULL
```

```
## Models are identical
```

```
## NULL
```

```
## Model selection table
```

```
##      (Int)  per_cpt corstr  qLik    QIC delta weight
## mmr_ex  88.25 0.0004735 exchn -1500 3007.5  0.00  0.969
## mmr_ar1 90.61 0.0003315   ar1 -1500 3015.7  8.25  0.016
## mmr_ind 90.61 0.0003313 indpnd -1500 3015.8  8.26  0.016
## Abbreviations:
## corstr: exchn = 'exchangeable', indpnd = 'independence'
## Models ranked by QIC(x)
```

1. ANOVA suggests that the three unadjusted models are all identical.
2. Exchangeable Structure has the lowest QIC.

Statistical Metrics:

- qLik: Quasi-likelihood of the model, a measure of the model fit to the data. Higher values (closer to zero, since these are likely negative) generally indicate a better fit.
- QIC: Quasi Information Criterion, a measure used for model selection. Lower QIC values indicate a better model fit considering both the goodness of fit and the complexity of the model.
- Delta: This column shows the difference in QIC from the best model to the current model. A delta of 0 indicates the best model.
- Weight: These are the Akaike weights based on the QIC values, providing a measure of each model's relative likelihood of being the best model among the set of models evaluated. A higher weight suggests a higher likelihood that this model is the best among those tested.

Adjusted GEE

```
## Models are identical

## NULL

## Models are identical

## NULL

## Models are identical

## NULL

## Model selection table
##      (Int)      n_sch  per_cpt typ corstr  qLik    QIC delta weight
## mmr_adj_ex  82.05 0.0003537 0.0004520  + exchnng -1500 3027.8  0.00      1
## mmr_adj_ar1 82.07 0.0005128 0.0004525  +   ar1 -1500 3046.3 18.54      0
## mmr_adj_ind 82.55 0.0005169 0.0004607  + indpnd -1500 3046.6 18.86      0
## Abbreviations:
## corstr: exchnng = 'exchangeable', indpnd = 'independence'
## Models ranked by QIC(x)

## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
## resized.
```

Table 7: Model Selection Result for MMR Rate

Structure	QIC	Delta	Weight
Unadjusted			
Exchangeable	3007.5	0.00	0.969

AR1	3015.7	8.25	0.016
Independence	3015.8	8.26	0.016
Adjusted			
Exchangeable	3027.8	0.00	1.000
AR1	3046.3	18.54	0.000
Independence	3046.6	18.86	0.000

```
## Analysis of 'Wald statistic' Table
##
## Model 1 mmr ~ per_capita + n_school + type
## Model 2 mmr ~ per_capita
##   Df      X2 P(>|Chi|)
## 1   5 3162.9 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Call:
## geeglm(formula = mmr ~ per_capita + n_school + type, family = gaussian,
##        data = mmr_samples, id = stateID, corstr = "exchangeable")
##
## Coefficients:
##              Estimate Std.err      Wald Pr(>|W|)
## (Intercept) 8.205e+01 1.960e+00 1752.911 < 2e-16 ***
## per_capita  4.520e-04 1.420e-04   10.135 0.001455 **
## n_school    3.537e-04 2.896e-04    1.492 0.221978
## typeOthers  7.673e+00 5.744e-01  178.429 < 2e-16 ***
## typePrivate 2.921e+00 7.651e-01   14.576 0.000135 ***
## typePublic  7.466e+00 2.683e-01  774.152 < 2e-16 ***
## typeUnknown 5.374e+00 1.270e+00   17.895 2.33e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation structure = exchangeable
## Estimated Scale Parameters:
##
##              Estimate Std.err
## (Intercept)   76.81   11.57
## Link = identity
##
## Estimated Correlation Parameters:
##              Estimate Std.err
## alpha    0.04515 0.03874
## Number of clusters: 21 Maximum cluster size: 708
```

Overall Unadjusted

```
## Models are identical

## NULL

## Models are identical
```

```
## NULL
```

```
## Models are identical
```

```
## NULL
```

```
## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
## resized.
```

Table 8: Unadjusted Model Selection Result for Overall Rate

Structure	QIC	Delta	Weight
Exchangeable	2934	0.00	1
Independence	3012	77.92	0
AR1	3012	77.94	0

```
## Models are identical
```

```
## NULL
```

```
## Models are identical
```

```
## NULL
```

```
## Models are identical
```

```
## NULL
```

```
## Model selection table
##           (Int)    n_sch  per_cpt typ corstr  qLik  QIC delta weight
## adj_overall_ar1 79.06 0.001278 0.0008127  +   ar1 -1499 3062  0.00  0.462
## adj_overall_ind 79.06 0.001277 0.0008108  + indpnd -1499 3062  0.05  0.451
## adj_overall_ex  80.56 0.001217 0.0006859  + exchnng -1499 3066  3.34  0.087
## Abbreviations:
## corstr: exchnng = 'exchangeable', indpnd = 'independence'
## Models ranked by QIC(x)
```

```
## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
## resized.
```

Table 9: Adjusted Model Selection Result for Overall Rate

Structure	QIC	Delta	Weight
AR1	3062	0.00	0.462
Independence	3062	0.05	0.451
Exchangeable	3066	3.34	0.087

```
## Analysis of 'Wald statistic' Table
##
## Model 1 overall ~ per_capita + n_school + type
## Model 2 overall ~ per_capita
##   Df    X2 P(>|Chi|)
## 1   4 1143   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##               Estimate   Std.err   Wald Pr(>|W|)
## (Intercept) 79.0551634 4.3637823 328.1968 0.000000
## per_capita  0.0008127 0.0004956   2.6883 0.101087
## n_school    0.0012781 0.0001059 145.7290 0.000000
## typePrivate -4.1125583 1.3244171   9.6422 0.001902
## typePublic  -2.0356892 1.2598845   2.6107 0.106143
## typeUnknown -2.9574209 3.9546106   0.5593 0.454556

## Warning in styling_latex_scale(out, table_info, "down"): Longtable cannot be
## resized.
```

Table 10: Best Model Estimation Result for MMR Vaccination Rate

X.	Estimate	Std.err	Wald	P.value
Intercept	79.0600	4.3640	328.197	< 0.05
Per Capita	0.0008	0.0005	2.688	0.101
Number of Schools	0.0013	0.0001	145.729	<0.05
School Type				
Private	-4.1100	1.3240	9.642	<0.05
Public	-2.0400	1.2600	2.611	0.106
Unknown	-2.9600	3.9550	0.559	0.454