University of Minnesota Morris Digital Well
University of Minnesota Morris Digital Well

Undergraduate Research Symposium 2021 | Undergraduate Research Symposium

4-14-2021

Conjunction of Factors Impacting the 2019-2020 Flu Season in the US

Yichen Wang

Follow this and additional works at: https://digitalcommons.morris.umn.edu/urs_2021

 Part of the Applied Statistics Commons, and the Influenza Humans Commons

Conjunction of Factors Impacting The 2019-2020 Flu Season in the US

Yichen Wang

Supported by the University of Minnesota's Office of Undergraduate Research

April 14th, 2021

Introduction

Literature Review

- Pica and Bouvier: influenza viruses spread more efficiently at relatively low relative humidity
- University of Chicago: high level of social connection promotes the spread of influenza.
- Dr. Jessica Grayson: babies and adults who are older than 65 are more likely to be contracted

My Goal

- To understand the conjugate impact of different aspects of variables on influenza cases
- To investigate the inner relationship between variables

Data Description

Variables

- Response variable: percentage of positive cases among all specimens (cdcfluview package)
- Predictors:
  - percentage of positive cases among all type A specimens
  - percentage of positive cases among all type B specimens (cdcfluview package)
  - average temperature
  - precipitation (NOAA's National Centers for Environmental Information)
  - relative humidity (rnoaa package - CO-OPS stations)

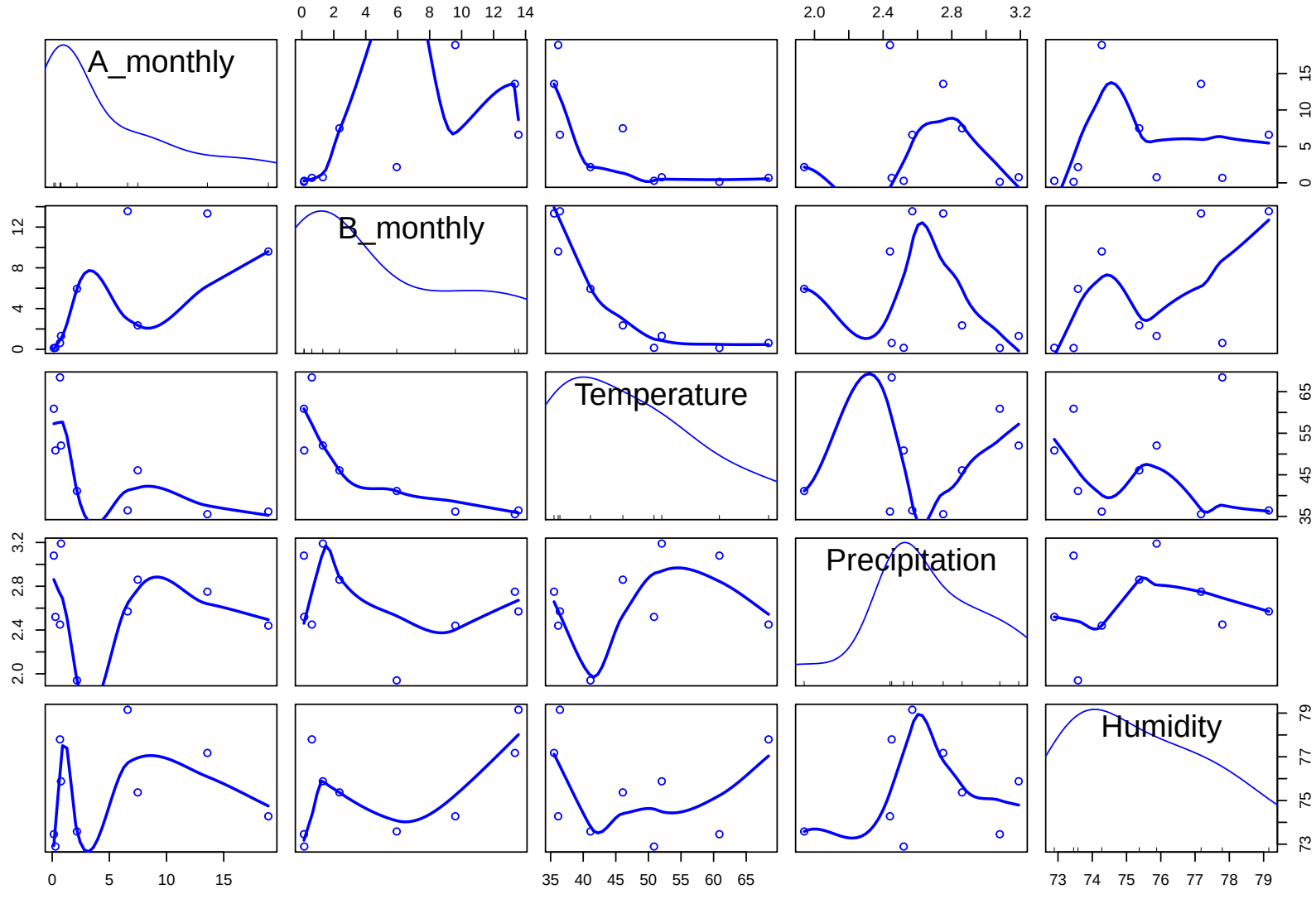
Variables

- Predictors continued:
  - vaccination coverage for 6 months to 17 years old
  - vaccination coverage for 18 to 49 years old
  - vaccination coverage for 50 to 64 years old
  - vaccination coverage for people older than 65 (CDC 2019-20 Influenza Season Vaccination Coverage Report)
  - population density (tidycensus package & United States by Density 2021 World Population Review)
- Other variables:
  - Date
  - State

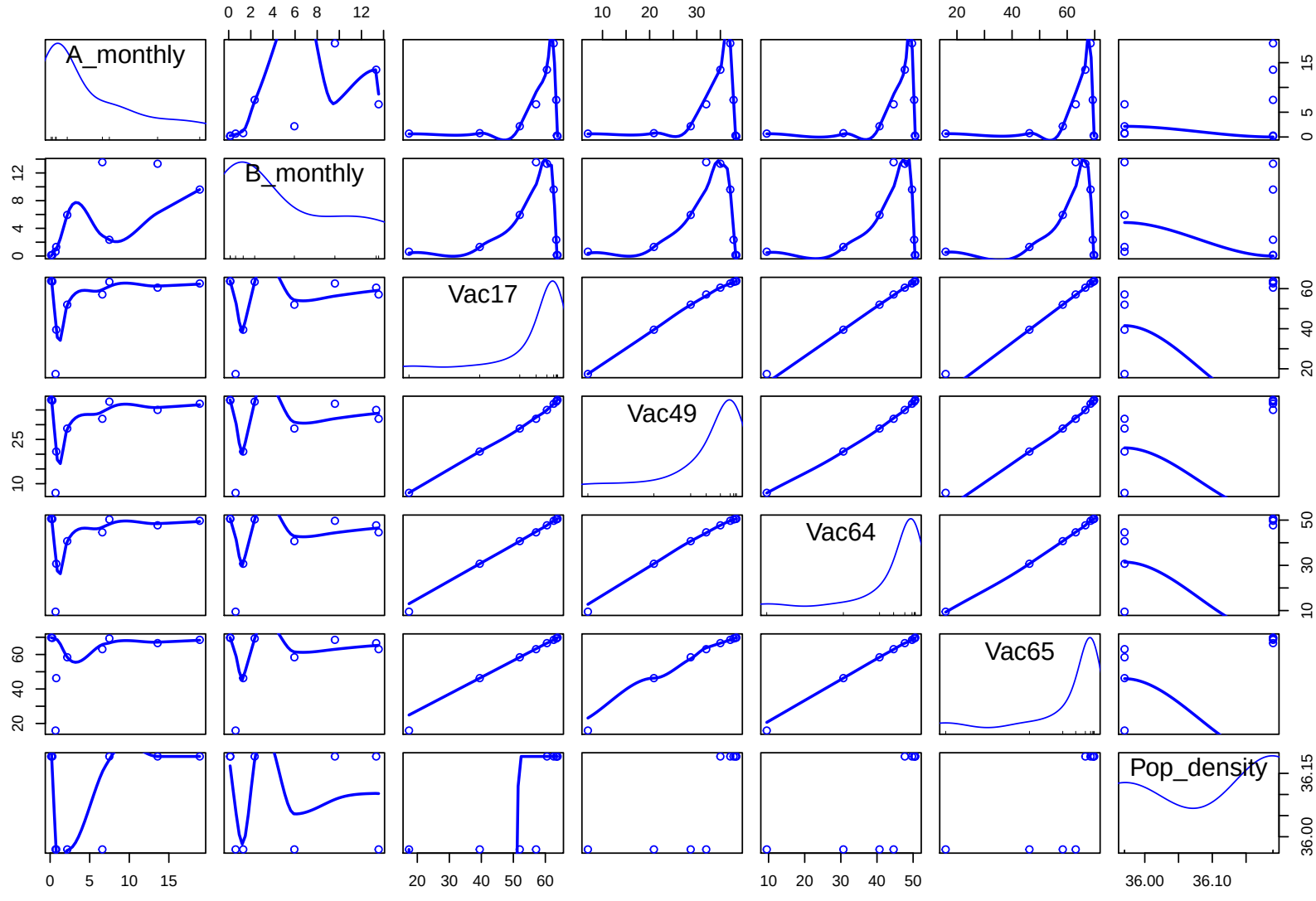
Data Frame

- variable name:
  - vaccination coverage for 6 months to 17 years old $\rightarrow$ Vac17
  - vaccination coverage for 18 to 49 years old $\rightarrow$ Vac49
  - vaccination coverage for 50 to 64 years old $\rightarrow$ Vac64
  - vaccination coverage for people older than 65 $\rightarrow$ Vac65
- monthly data - 9 months
- two data frames: National and State
  - National: 9 observations, 10 variables
  - State: 42 states, 9 variables, 441 observations

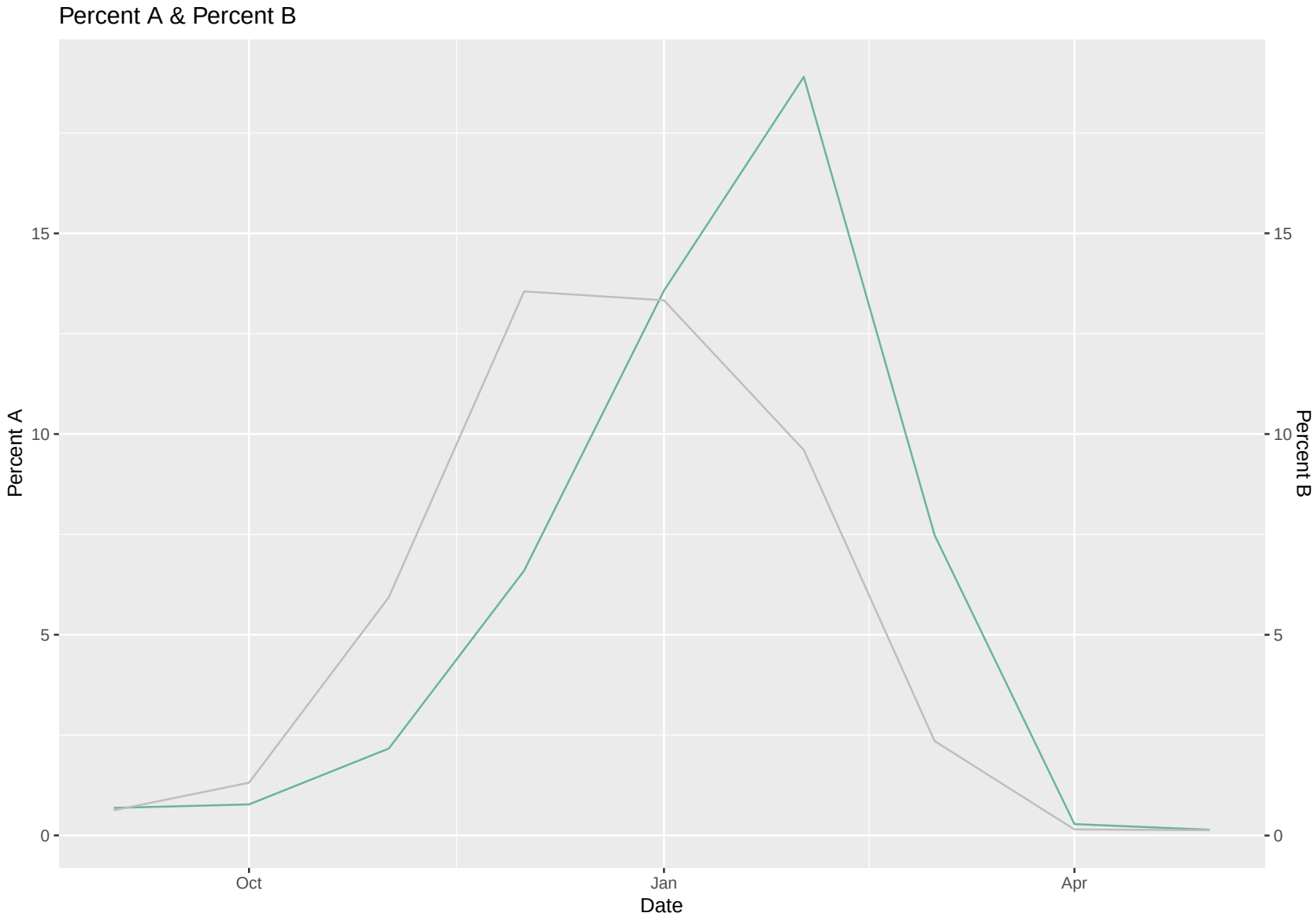
Basic Graphical Displays



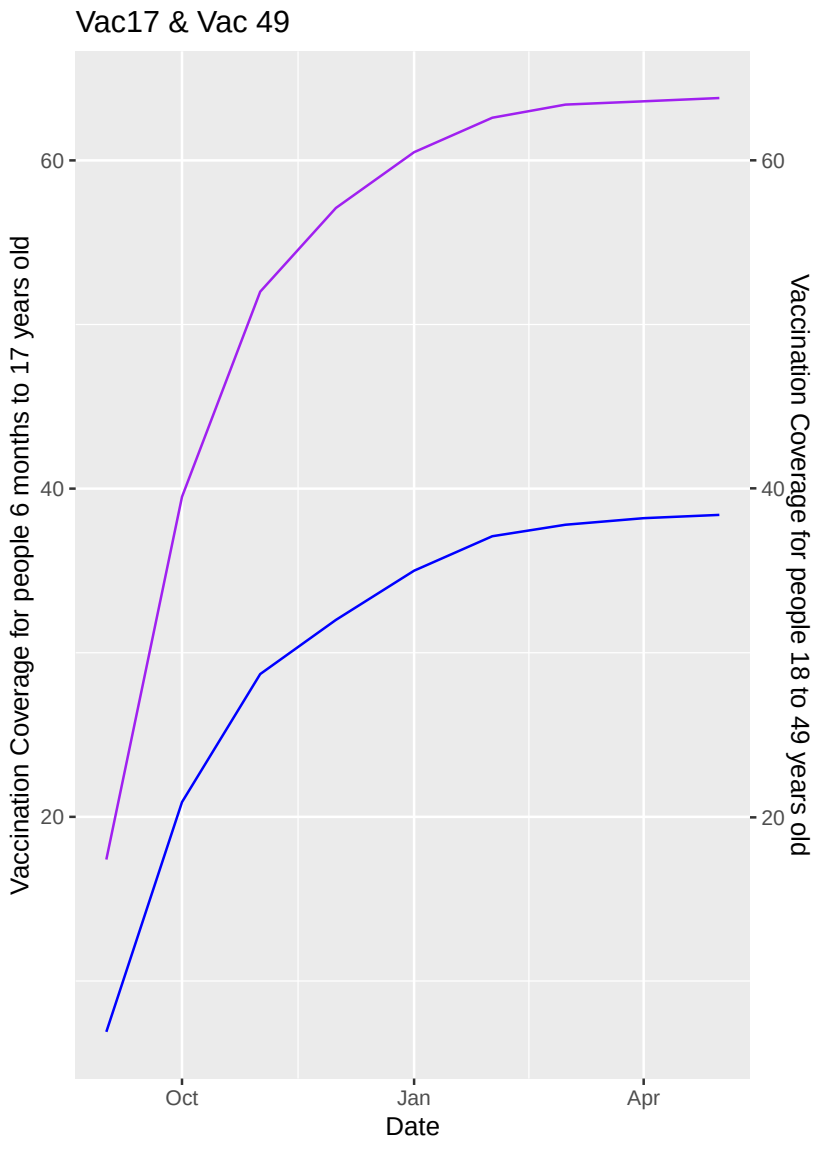
Basic Graphical Displays

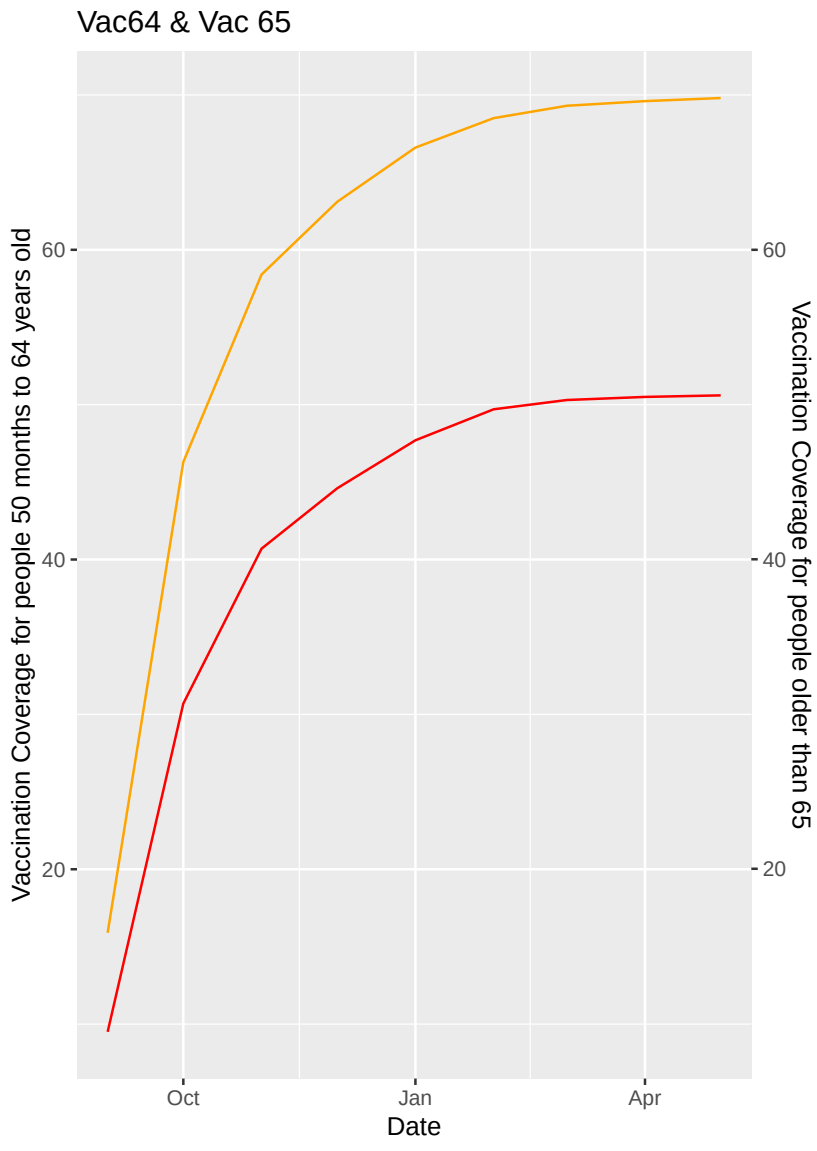


Basic Graphical Displays

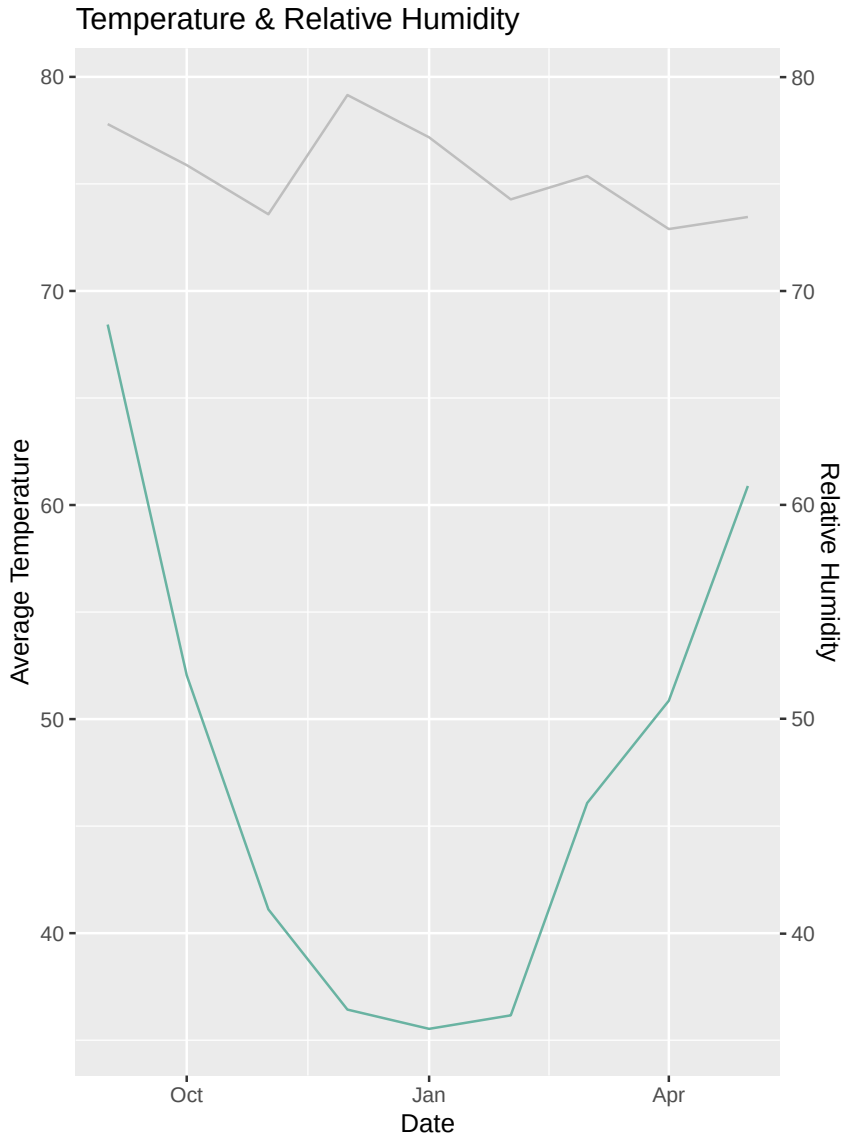


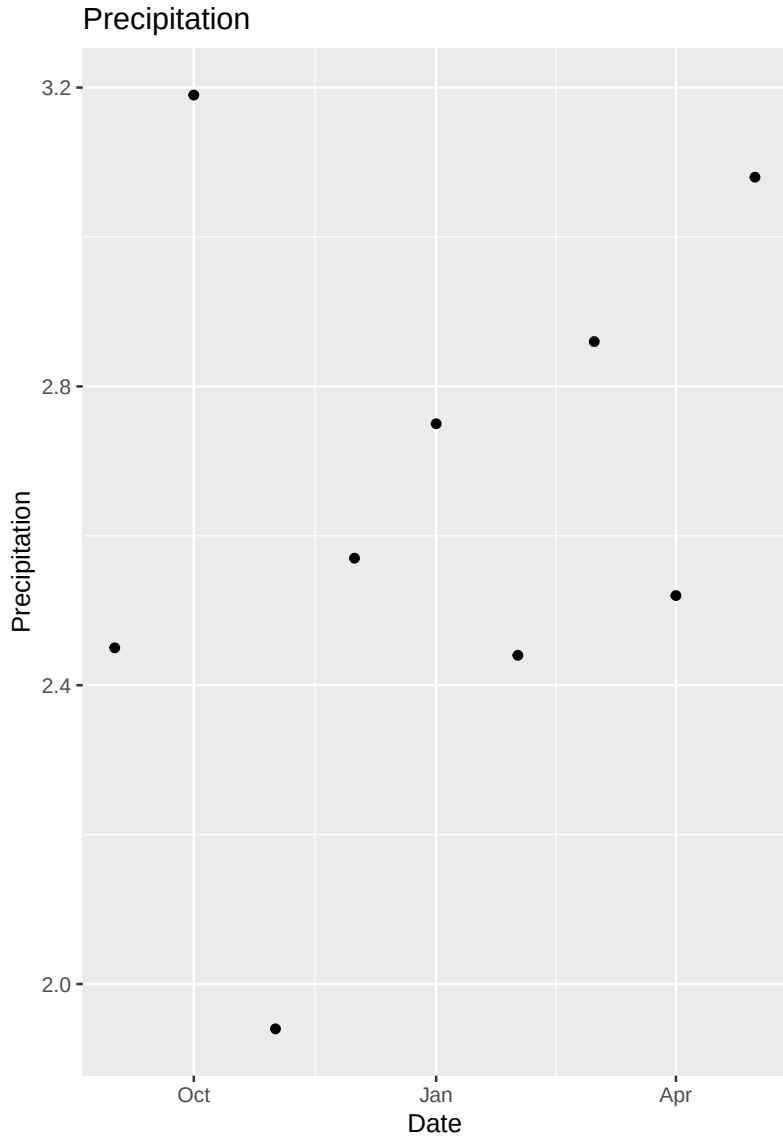
Basic Graphical Displays





Basic Graphical Displays





Statistical Methods & Results

Multiple Regression National Type A

```
modelA<-lm(A_monthly~Temperature+Vac17+Vac49,data=National[,c(-2:-1,-4)])
summary(modelA)
```

```
##
## Call:
## lm(formula = A_monthly ~ Temperature + Vac17 + Vac49, data = National[,
##     c(-2:-1, -4)])
##
## Residuals:
##      9     10     11     12     13     14     15     16     17
##  0.775 -1.212 -1.594  0.862  1.630  2.105 -0.741 -5.522  3.698
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  114.862     29.338    3.92    0.011 *
## Temperature   -0.985      0.216   -4.57    0.006 **
## Vac17         -6.240      1.962   -3.18    0.025 *
## Vac49          8.845      2.788    3.17    0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.4 on 5 degrees of freedom
## Multiple R-squared:  0.842,  Adjusted R-squared:  0.748
## F-statistic:  8.9 on 3 and 5 DF,  p-value: 0.019
```

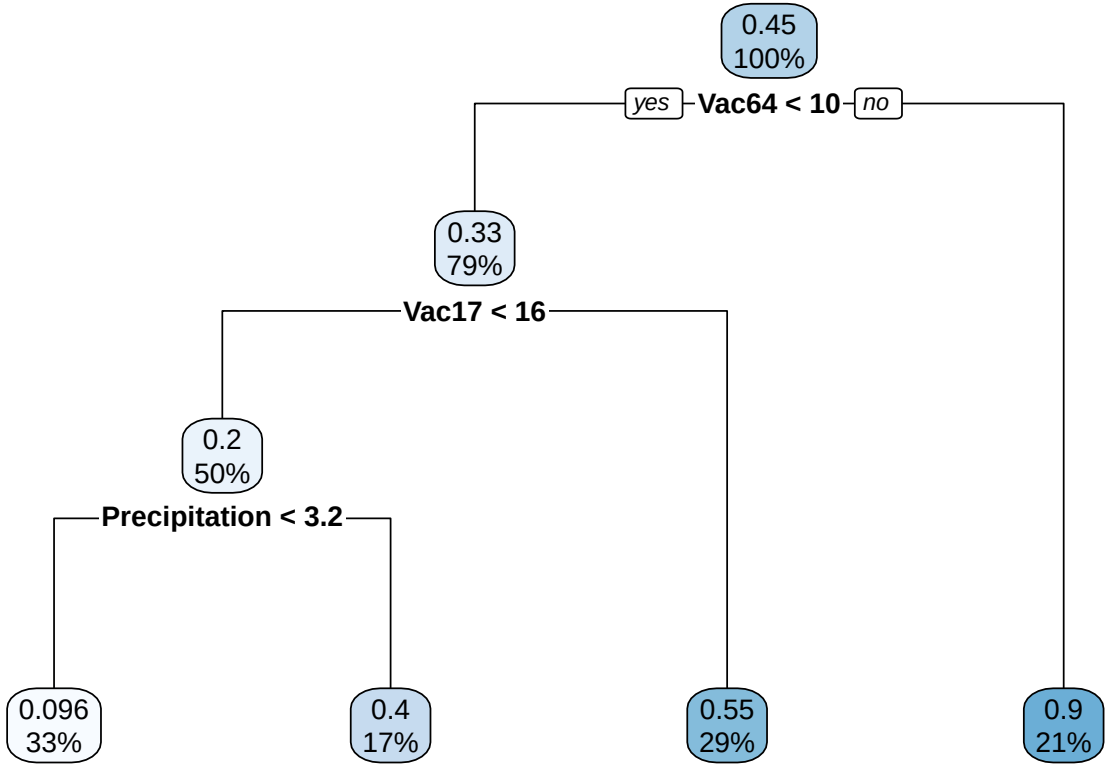
Multiple Regression National Type B

```r
modelB<-lm(B_monthly~Temperature+Humidity,data=National[,-3:-1])
summary(modelB)
```

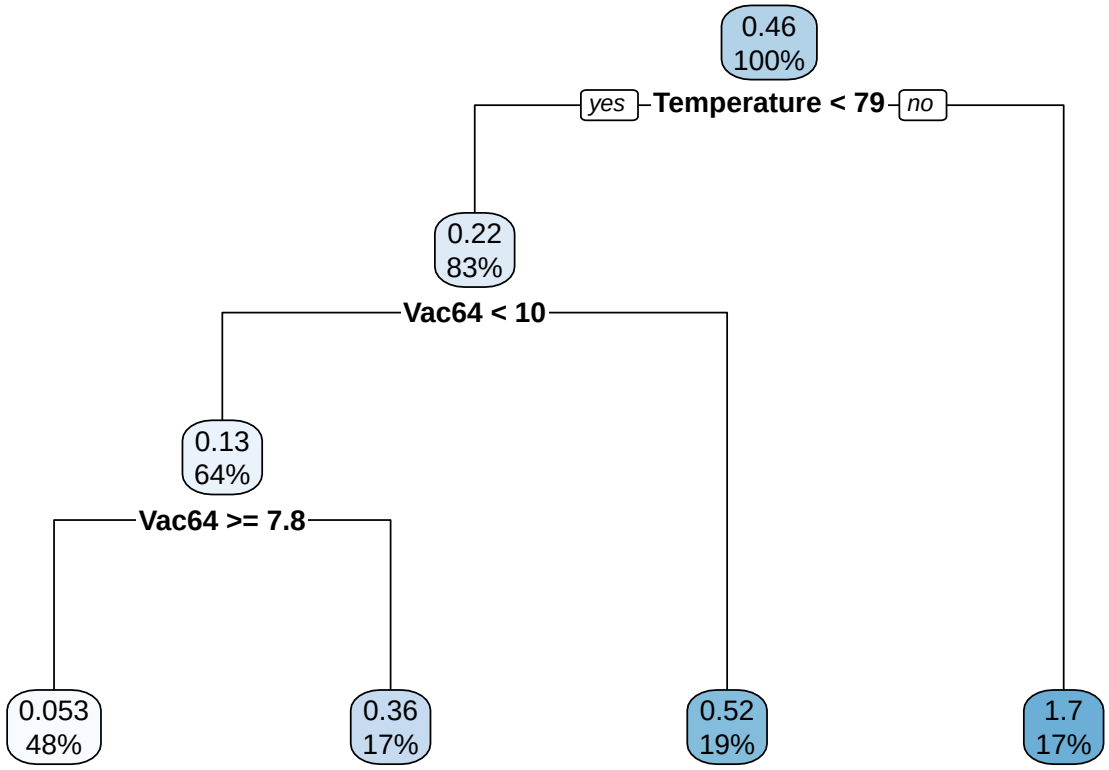
```
##
## Call:
## lm(formula = B_monthly ~ Temperature + Humidity, data = National[,
##     -3:-1])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.265 -0.915  0.379  1.388  2.288
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -59.4535    27.7460   -2.14   0.0759 .
## Temperature  -0.3830     0.0672   -5.70   0.0013 **
## Humidity      1.0975     0.3606    3.04   0.0227 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.2 on 6 degrees of freedom
## Multiple R-squared:  0.884,  Adjusted R-squared:  0.845
## F-statistic: 22.9 on 2 and 6 DF,  p-value: 0.00156
```

Regression Tree Statewide Type A

- to predict the values of a continuous response variable
- predictions are based on the importance of predictor variables, from the most important to the least important

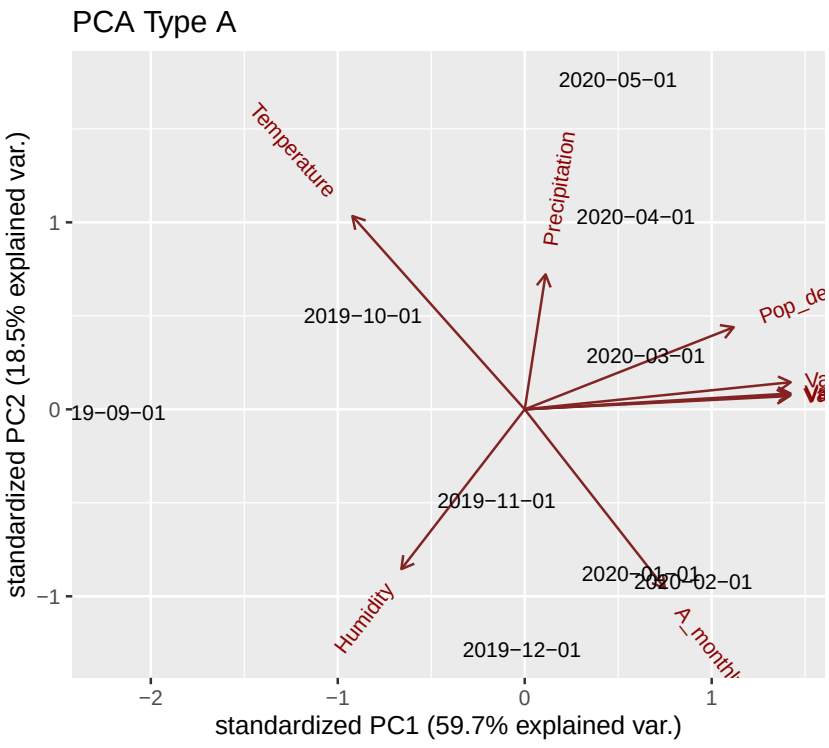


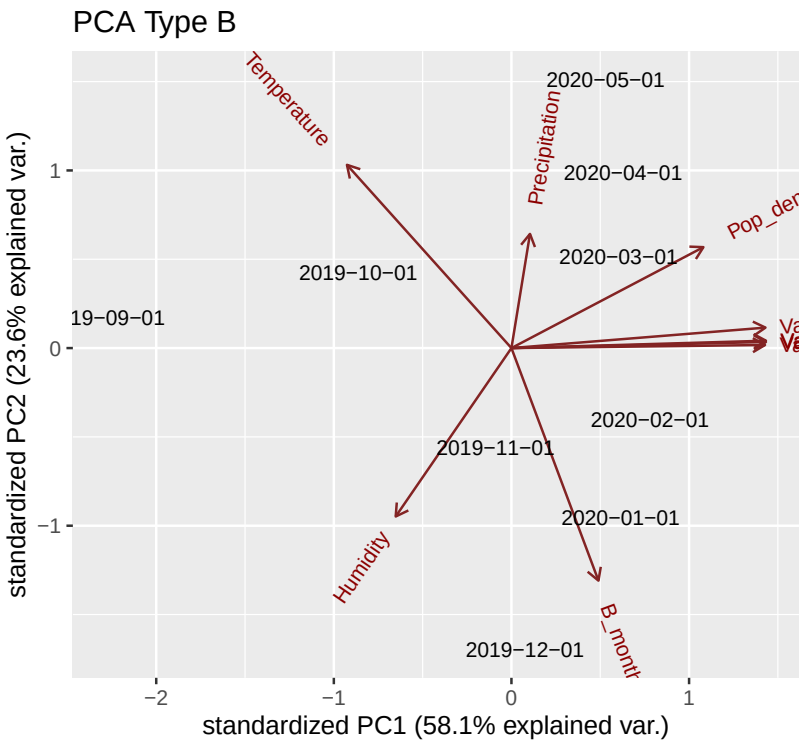
Regression Tree Statewide Type B



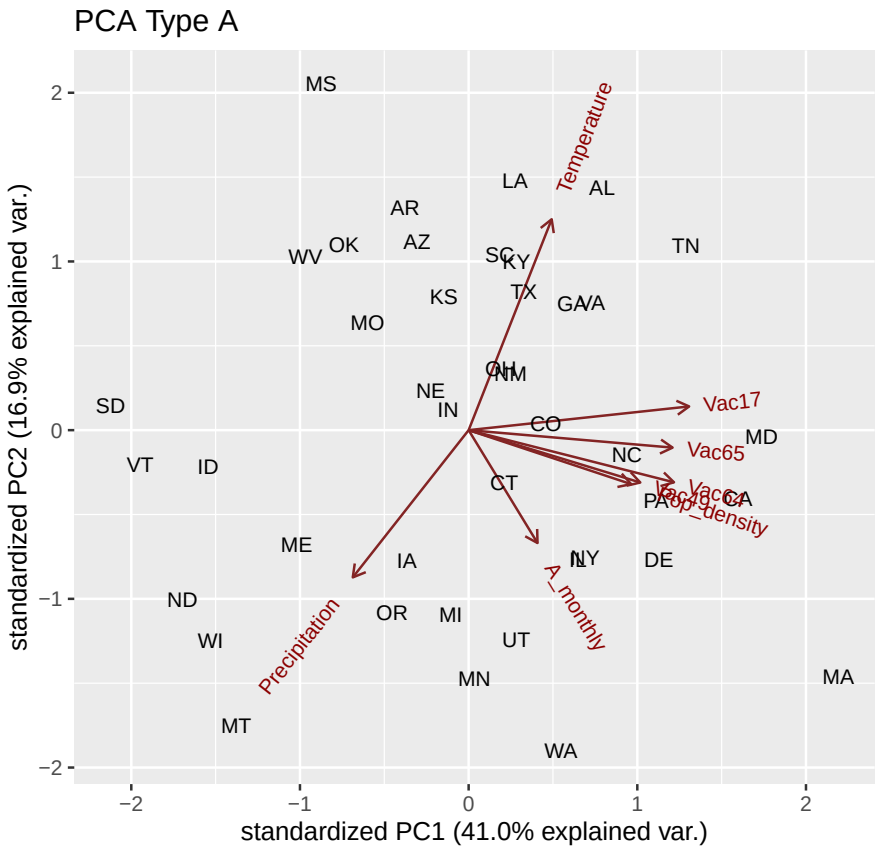
PCA (Principal Component Analysis) National

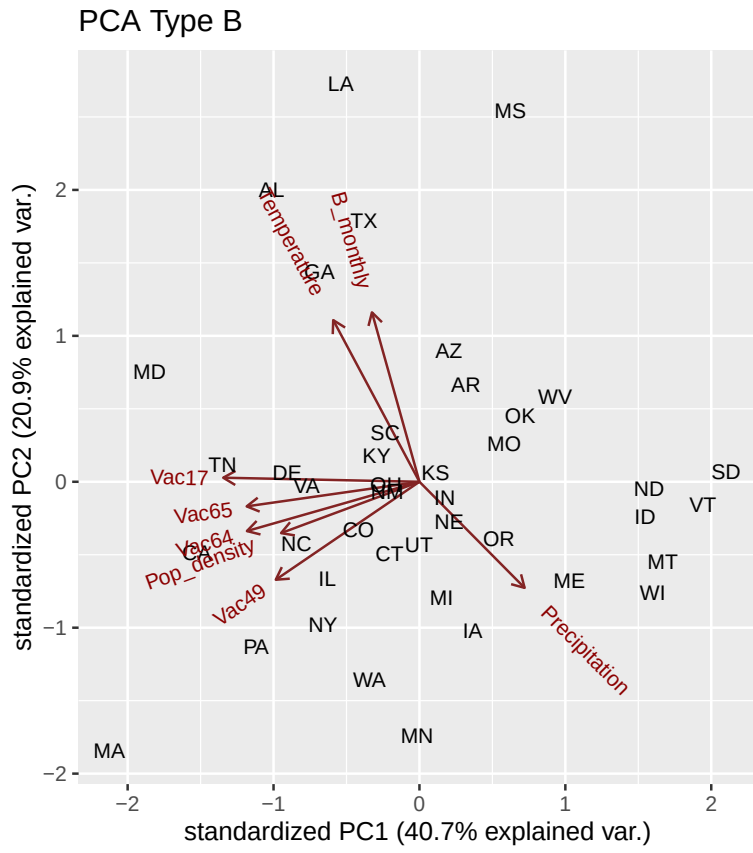
- data reduction method



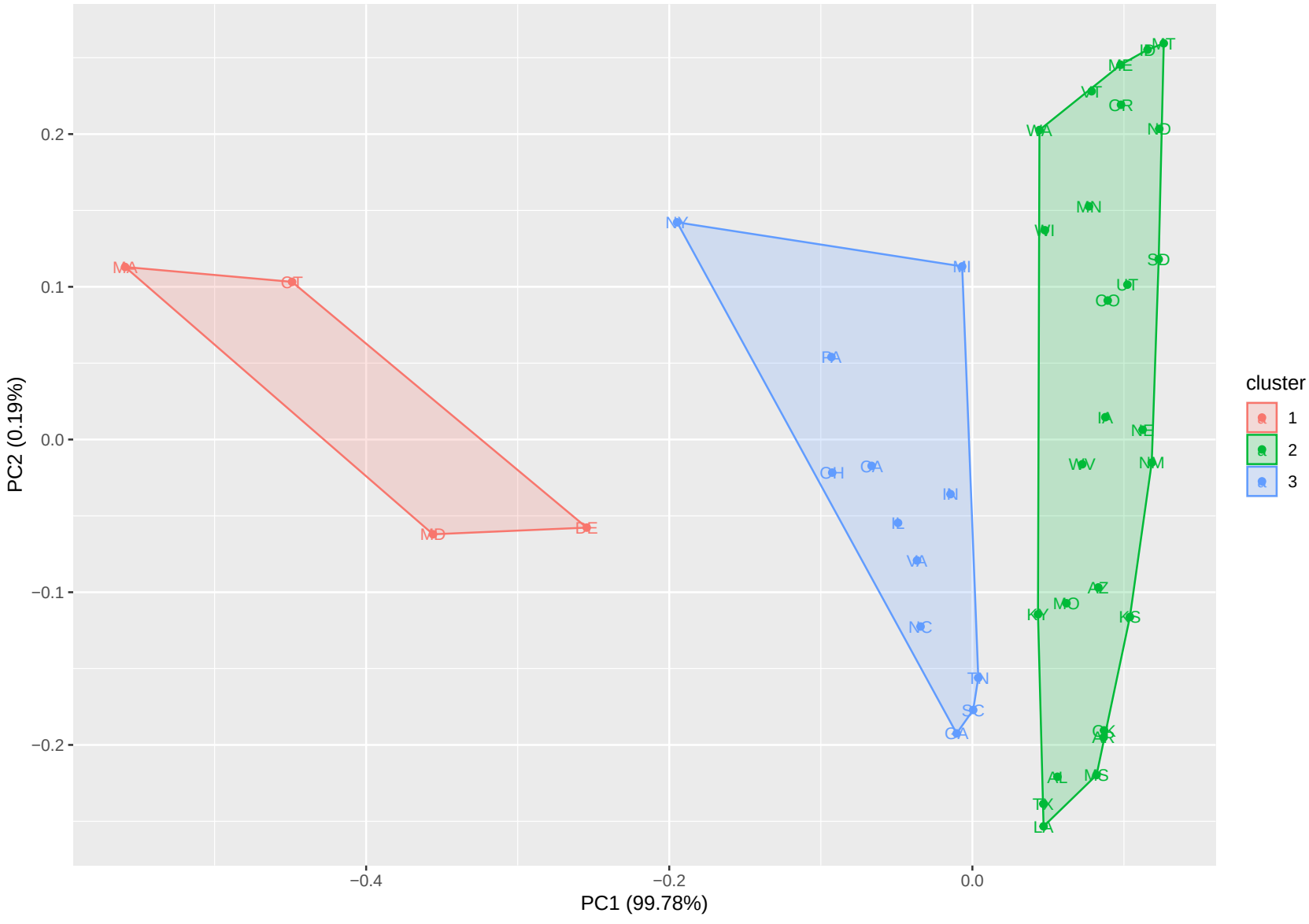


PCA Statewide





Cluster Analysis Statewide Type A

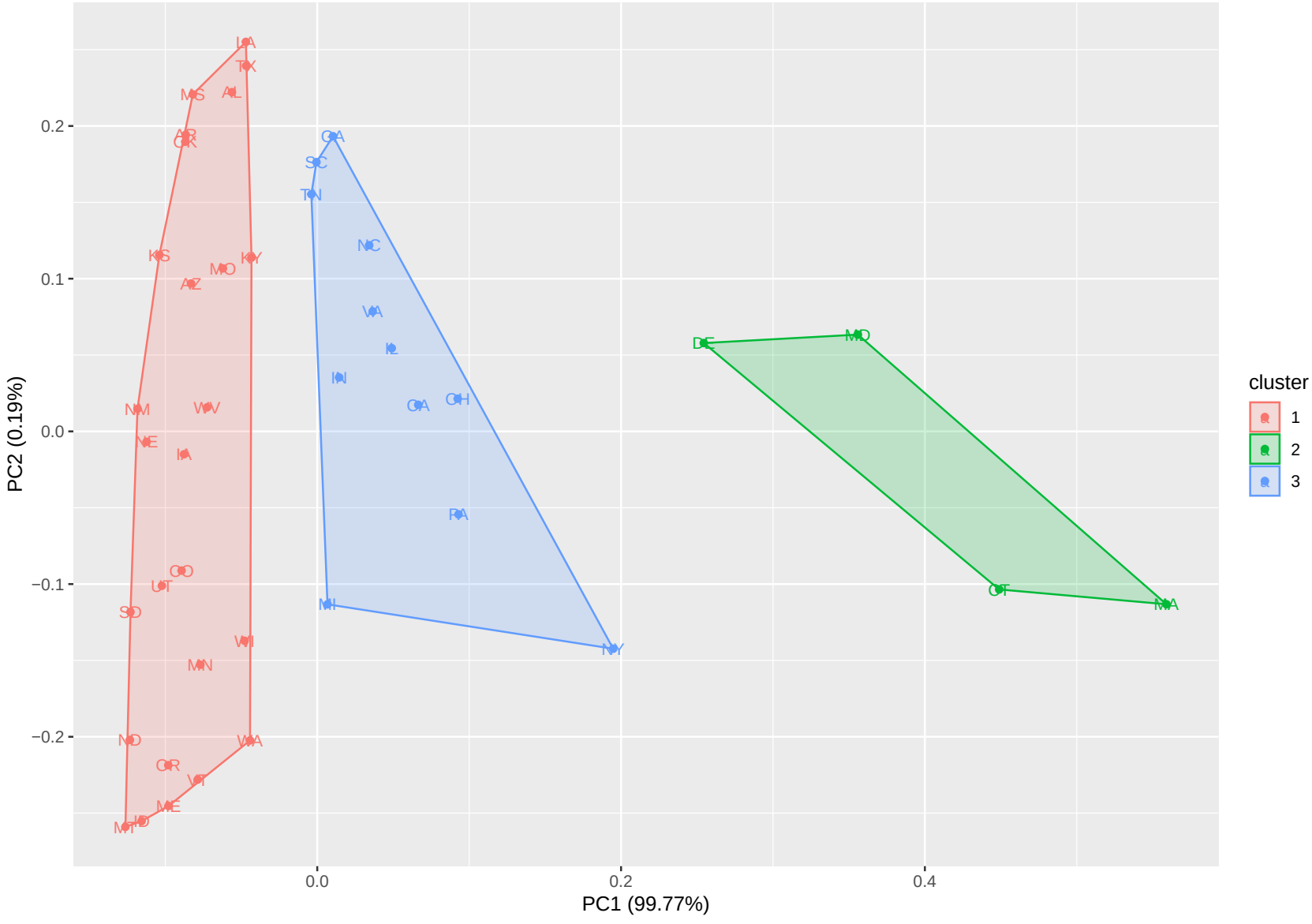


Cluster Differences

```
## [1] 26 12  4
```

```
##                  [,1] [,2]   [,3]
## A_monthly         0.4  0.4   0.89
## Temperature      70.8 67.8  67.05
## Precipitation     2.3  2.7   1.45
## Vac17            17.5 14.7  18.30
## Vac49             7.2  6.2   6.97
## Vac64             9.7  8.4  10.15
## Vac65            16.3 13.7  16.55
## Pop_density     230.6 59.6 681.96
```

Cluster Analysis Statewide Type B



Cluster differences

	[,1]	[,2]	[,3]
B_monthly	0.5108	0.2825	0.6875
Temperature	67.8462	70.8333	67.0500
Precipitation	2.7396	2.2792	1.4450
Vac17	14.6731	17.5000	18.3000
Vac49	6.2462	7.2083	6.9750
Vac64	8.4038	9.7250	10.1500
Vac65	13.6923	16.3083	16.5500
Pop_density	59.5643	230.5678	681.9558

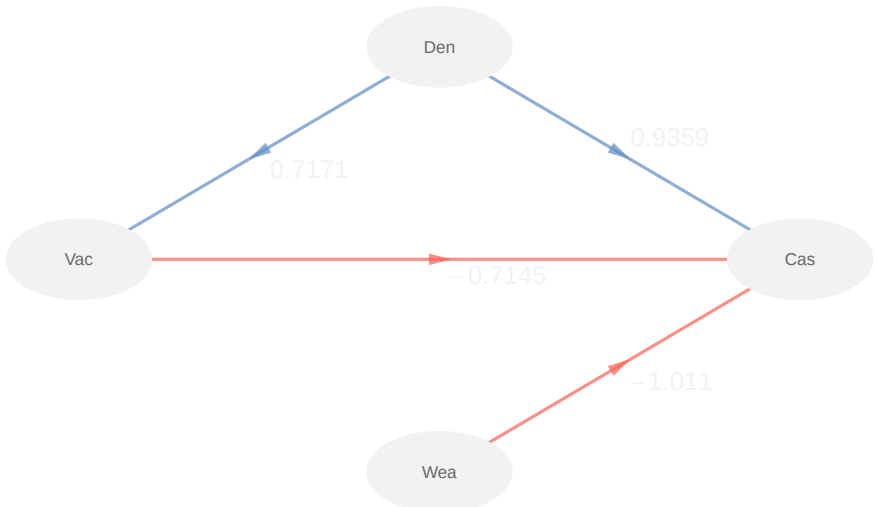
Path Analysis National Type A

- causal relationship between factors

```
$Vac
               Estimate Std. Error       t value   Pr(>|t|)
Intercept -1.020118e-14  0.2634469 -3.872194e-14 1.00000000
Den        7.170566e-01  0.2634469  2.721826e+00 0.02968717

$Cas
               Estimate Std. Error       t value    Pr(>|t|)
Intercept -1.276646e-14  0.1796786 -7.105165e-14 1.000000000
Den        9.359158e-01  0.2942918  3.180231e+00 0.024531871
Vac       -7.144872e-01  0.3267158 -2.186877e+00 0.080412602
Wea       -1.011196e+00  0.2277342 -4.440244e+00 0.006763333
```

Figure 1: Inner Plot Loadings Type A



Outer Plot

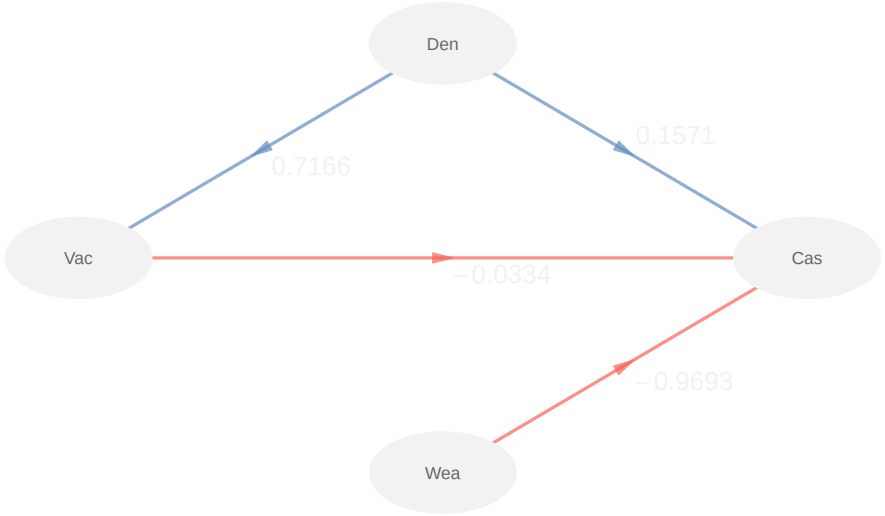
name	loading
Pop_density	1.0000
Vac17	0.9997
Vac49	0.9975
Vac64	0.9998
Vac65	0.9976
Temperature	0.9720
Precipitation	0.3897
Humidity	-0.2802
A_monthly	1.0000

Path Analysis National Type B

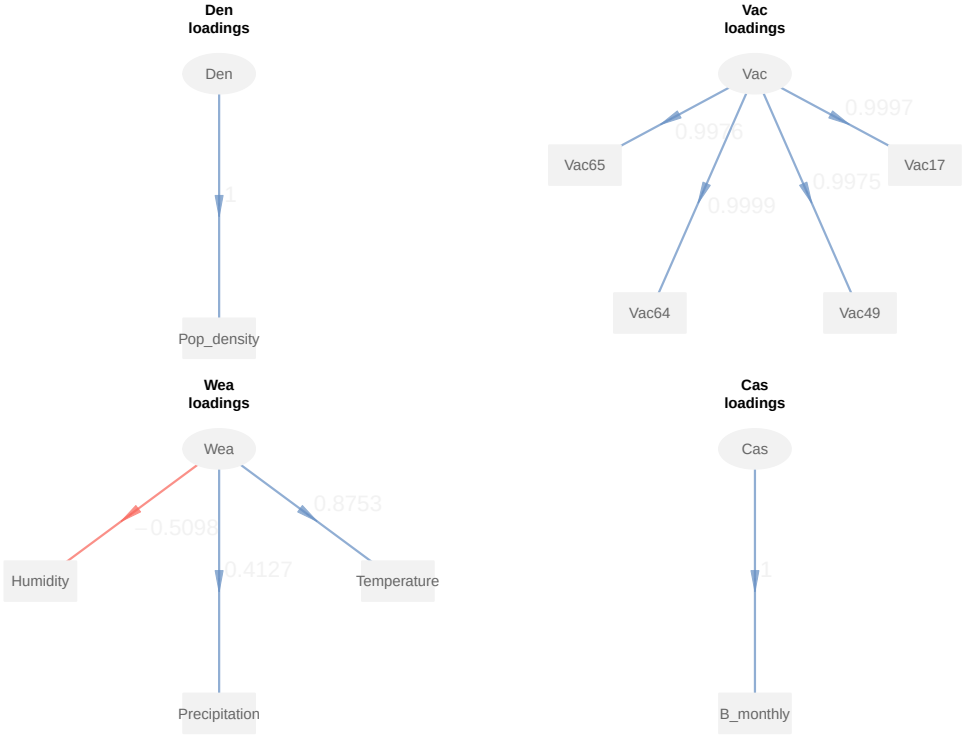
```
$Vac
              Estimate Std. Error       t value   Pr(>|t|)
Intercept -1.019418e-14  0.2636351 -3.866778e-14 1.00000000
Den        7.165722e-01  0.2636351  2.718046e+00 0.02984979

$Cas
              Estimate Std. Error       t value    Pr(>|t|)
Intercept -1.295574e-15  0.1453247 -8.915026e-15 1.000000000
Den        1.571337e-01  0.2360192  6.657667e-01 0.535038613
Vac       -3.340362e-02  0.2388510 -1.398513e-01 0.894236421
Wea       -9.692878e-01  0.1688070 -5.741989e+00 0.002245396
```

Figure 2: Inner Plot Loadings Type B



Outer Plot



name	loading
Pop_density	1.0000
Vac17	0.9997
Vac49	0.9975
Vac64	0.9999
Vac65	0.9976
Temperature	0.8753
Precipitation	0.4127
Humidity	-0.5098
B_monthly	1.0000

Conclusion

Conclusion

1. temperature is negatively related to both Type A and Type B
2. vaccination coverage is related to Type A but not Type B
3. vaccination coverage for different age groups are closely related
4. population density affects vaccination coverage

Reference

Reference

1. Pica, N., and Bouvier, N. M. (2012), "Environmental Factors Affecting the Transmission of Respiratory Viruses," Current Opinion in Virology, 2, 90–95. https://doi.org/10.1016/j.coviro.2011.12.003.

2. "Data analysis shows what drives the spread of flu" (n.d.). Available at https://www.uchicagomedicine.org/forefront/biological-sciences-articles/massive-data-analysis-shows-what-drives-the-spread-of-flu

3. News, A. B. C. (n.d.). "Flu shot better than last year, despite tough season for kids," ABC News, Available at https://abcnews.go.com/Health/1300-people-died-flu-year/story?id=67754182.