Model_v1

Zi Yang

2021/4/7

Process county death data

```
Sys.setlocale("LC_TIME", "English")
```

```
## [1] "English_United States.1252"
```

```
library(readxl)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(dplyr)
library(ggplot2)
library(ggpubr)
library(cowplot)
```

```
##
## Attaching package: 'cowplot'
```

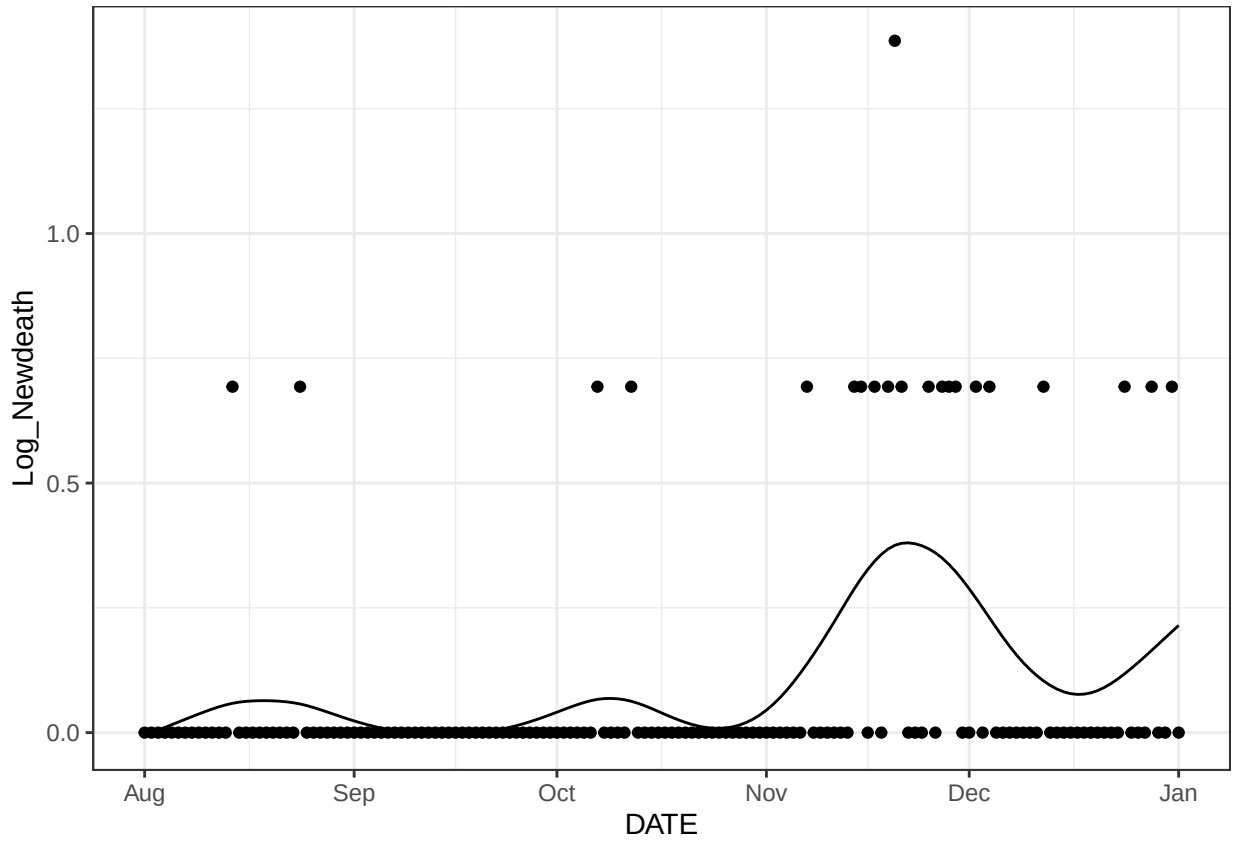
```
## The following object is masked from 'package:ggpubr':
##
##     get_legend
```

```
covid<-read_xlsx("D:/Study/CMU/Statistical Practice/EDA/bmodel/COVID_CASES_OH_CNTY_20210223_pop.xlsx")
covid$DATE<-as.Date(covid$DATE,"%m/%d/%Y")
death<-covid%>%
  filter(DATE<=as.Date("2021-01-01")
                      & DATE>=as.Date("2020-08-01"))%>%
  select(COUNTY,DATE,NEWDEATHS,CUMDEATHS)%>%
  mutate(Log_Newdeath=log(NEWDEATHS+1),
         Log_Cumdeath=log(CUMDEATHS+1))

county_name<-as.vector(distinct(death,COUNTY)[-c(63,81),])
```

ADAMS County Test

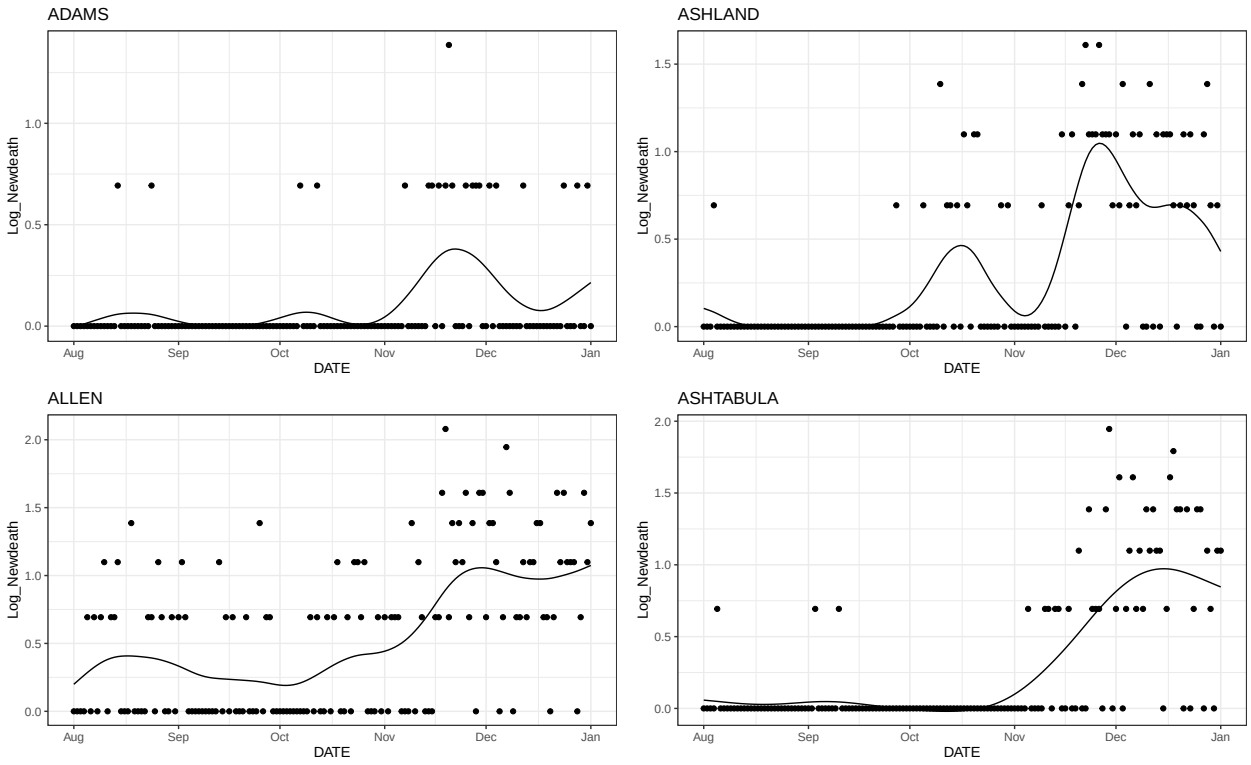
```
allen<-death%>%filter(COUNTY=="ADAMS")
spline.allen<-smooth.spline(x = allen$DATE, y = allen$Log_Newdeath)
ggplot(data=allen)+geom_point(aes(x=DATE,y=Log_Newdeath))+
  geom_line(aes(x=as.Date(spline.allen$x,origin="1970-01-01"),y=spline.allen$y))+
  theme_bw()
```

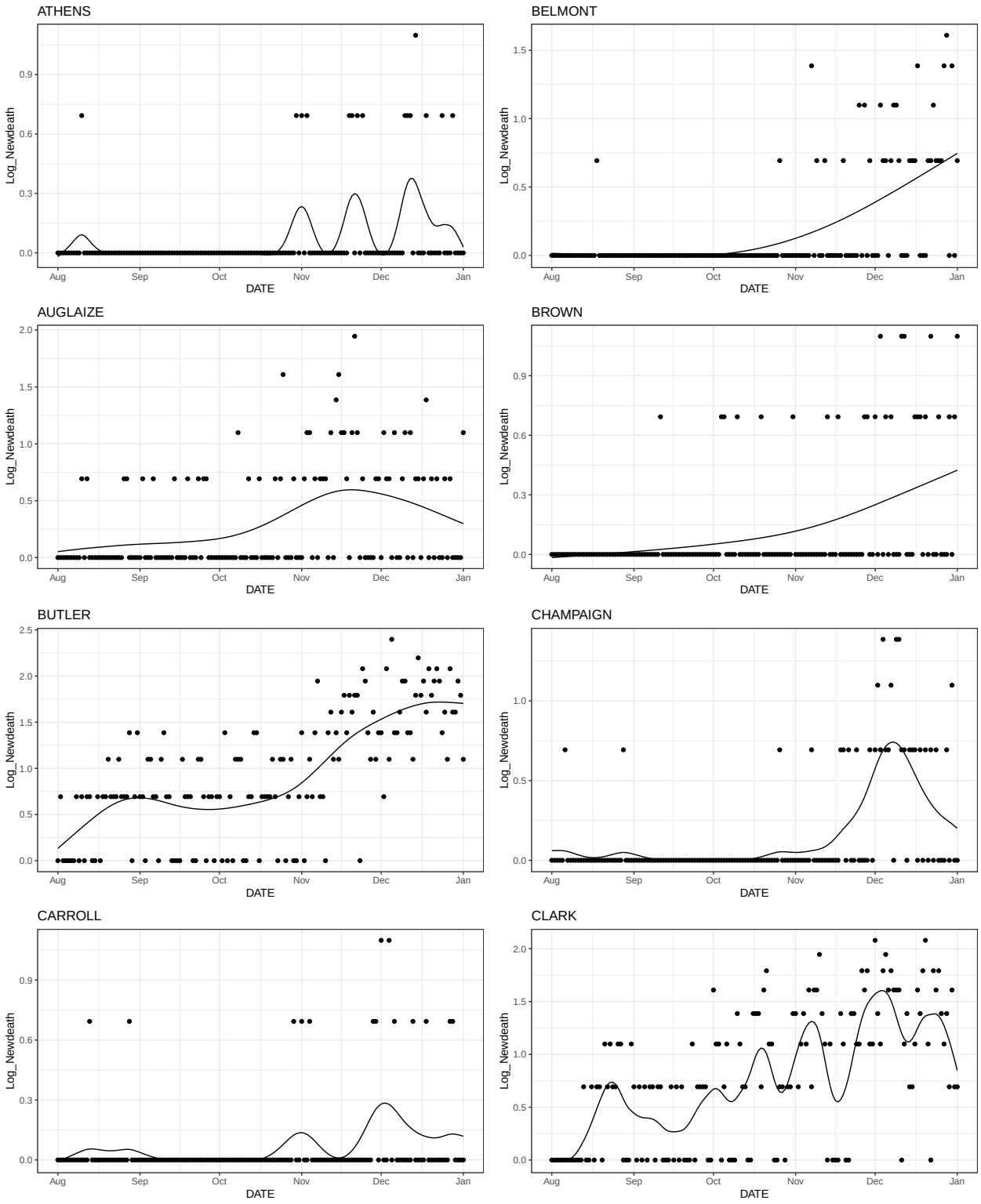


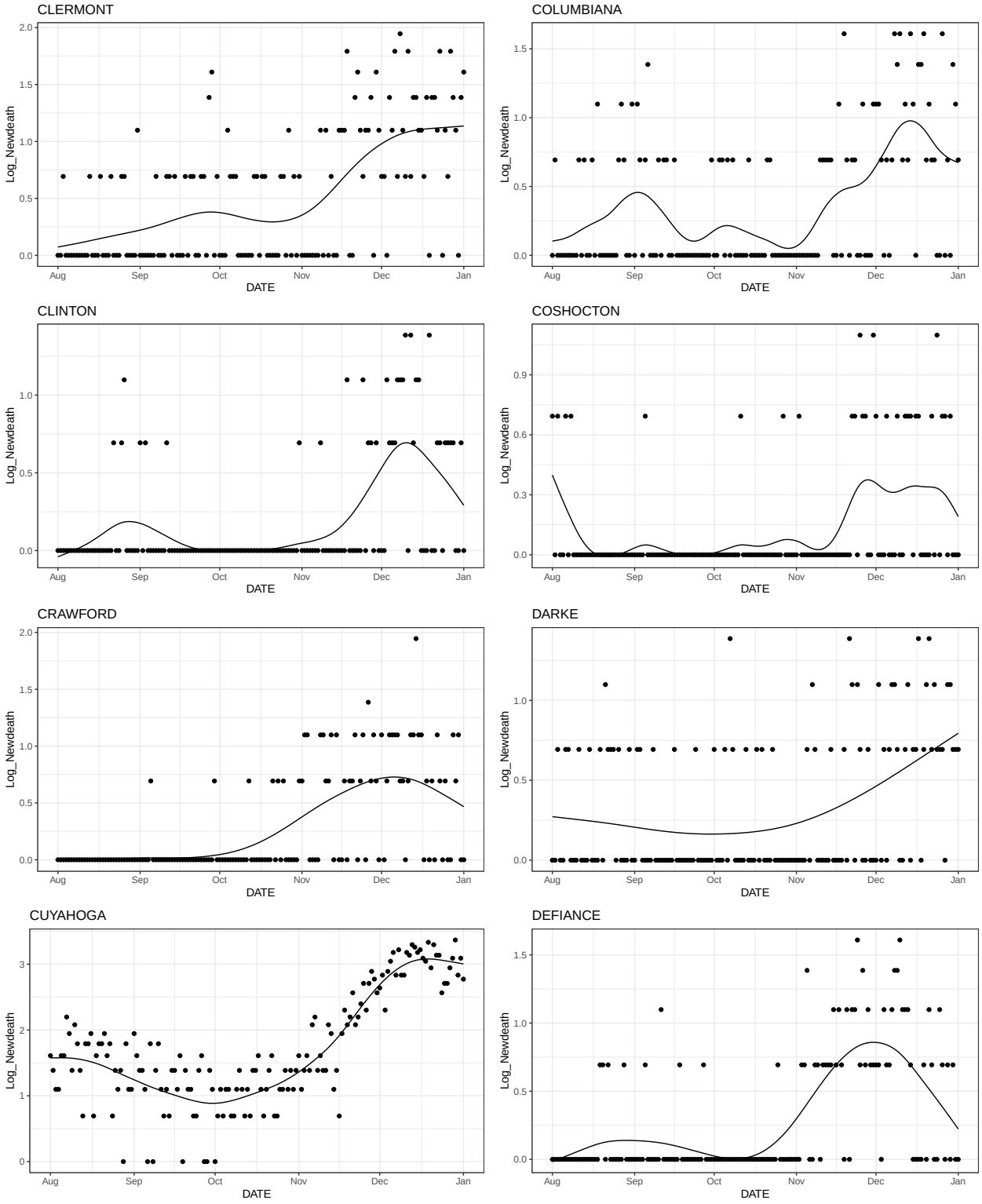
County data and spline

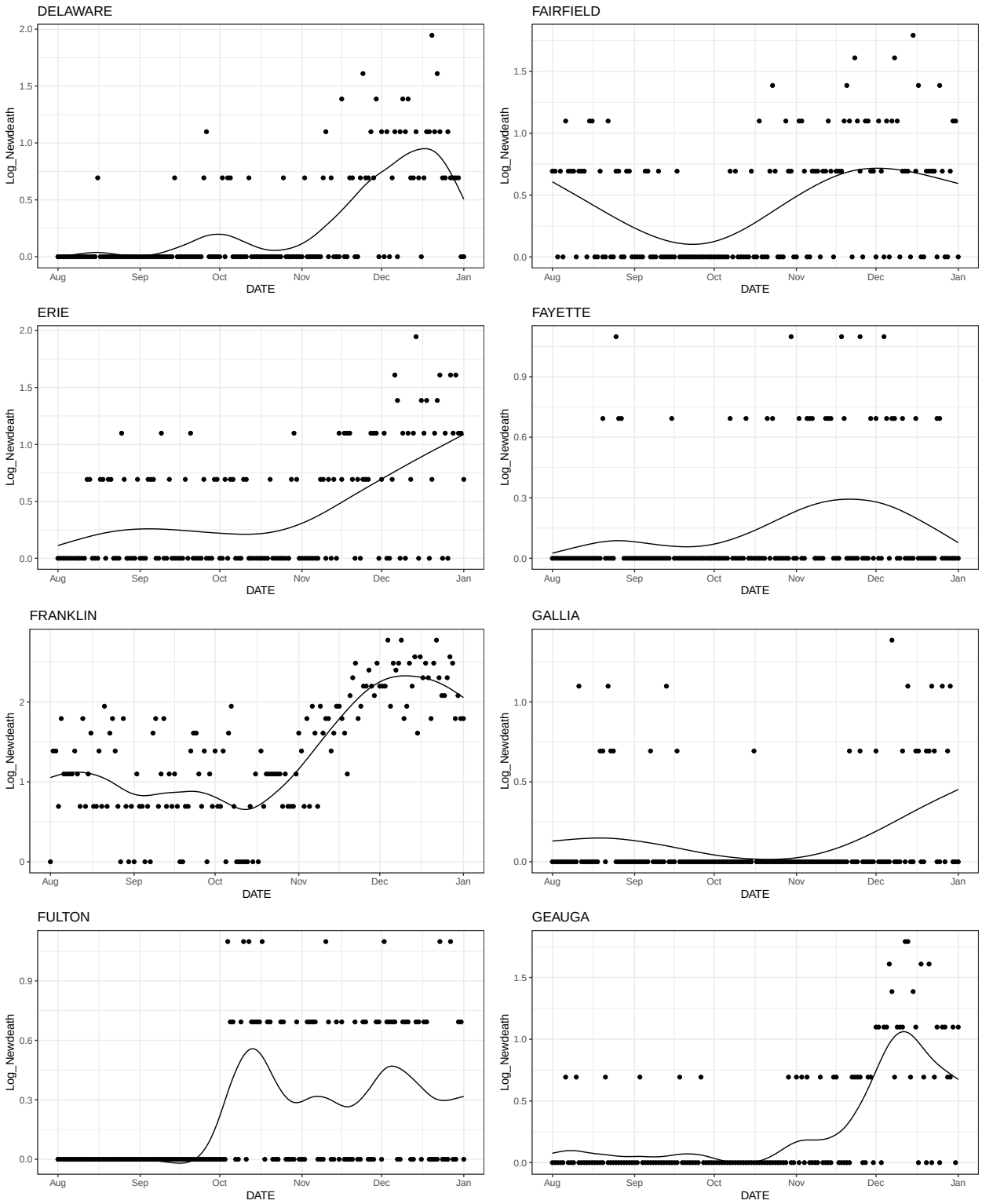
```
for(i in 1:22){
  county1<-death%>%filter(COUNTY==county_name$COUNTY[(i-1)*4+1])
  county2<-death%>%filter(COUNTY==county_name$COUNTY[(i-1)*4+2])
  county3<-death%>%filter(COUNTY==county_name$COUNTY[(i-1)*4+3])
  county4<-death%>%filter(COUNTY==county_name$COUNTY[(i-1)*4+4])
  spline1<-smooth.spline(x = county1$DATE, y = county1$Log_Newdeath)
  spline2<-smooth.spline(x = county2$DATE, y = county2$Log_Newdeath)
  spline3<-smooth.spline(x = county3$DATE, y = county3$Log_Newdeath)
  spline4<-smooth.spline(x = county4$DATE, y = county4$Log_Newdeath)
  p1<-ggplot(data=county1)+geom_point(aes(x=DATE,y=Log_Newdeath))+
  geom_line(aes(x=as.Date(spline1$x,origin="1970-01-01"),y=spline1$y))+
  theme_bw()+labs(title=county_name$COUNTY[(i-1)*4+1])
  p2<-ggplot(data=county2)+geom_point(aes(x=DATE,y=Log_Newdeath))+
  geom_line(aes(x=as.Date(spline2$x,origin="1970-01-01"),y=spline2$y))+
  theme_bw()+labs(title=county_name$COUNTY[(i-1)*4+2])
  p3<-ggplot(data=county3)+geom_point(aes(x=DATE,y=Log_Newdeath))+
```

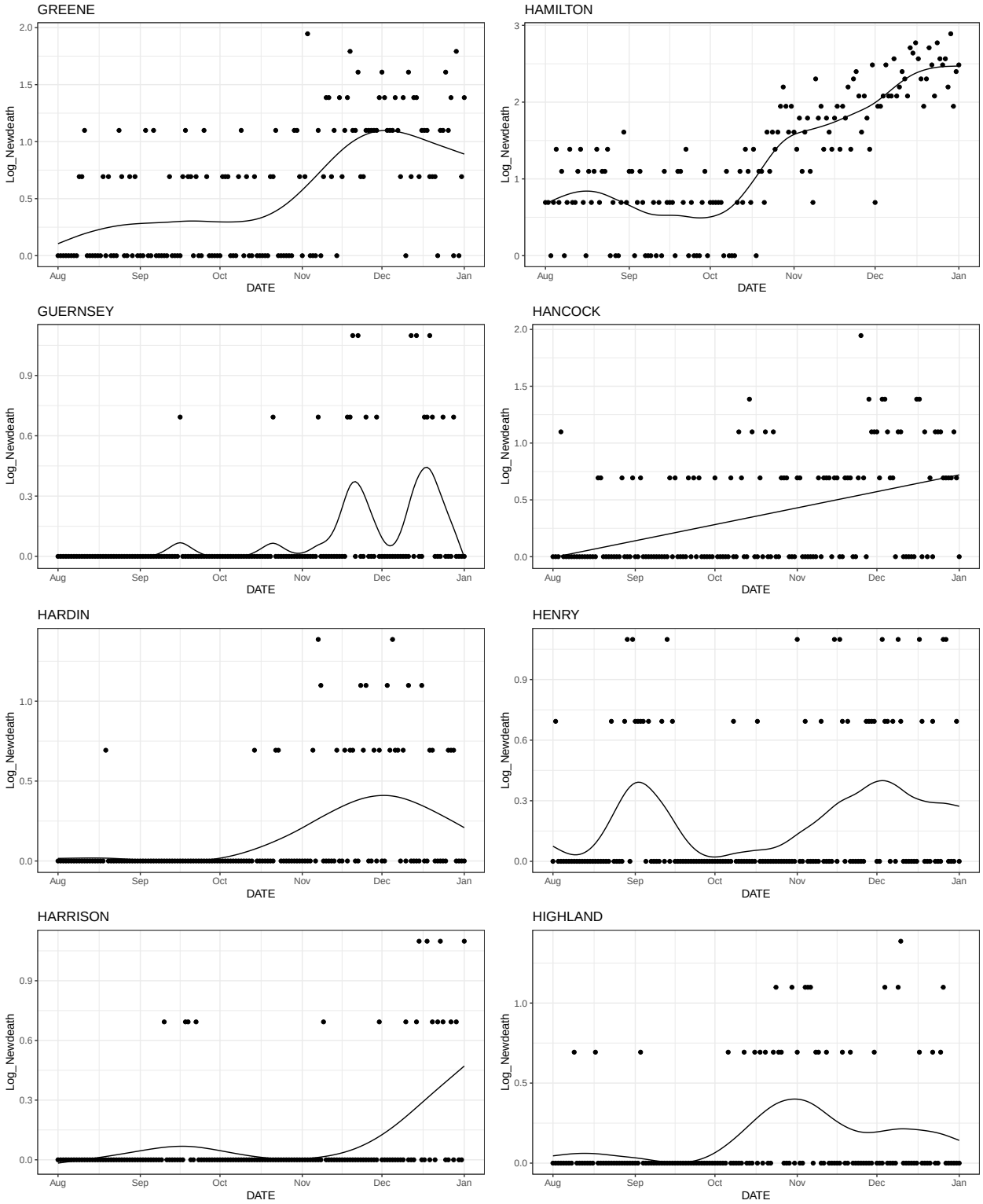
```
  geom_line(aes(x=as.Date(spline3$x,origin="1970-01-01"),y=spline3$y))+
  theme_bw()+labs(title=county_name$COUNTY[(i-1)*4+3])
  p4<-ggplot(data=county4)+geom_point(aes(x=DATE,y=Log_Newdeath))+
  geom_line(aes(x=as.Date(spline4$x,origin="1970-01-01"),y=spline4$y))+
  theme_bw()+labs(title=county_name$COUNTY[(i-1)*4+4])
  print(plot_grid(p1, p2,ncol = 1, nrow = 2))
  print(plot_grid(p3, p4,ncol = 1, nrow = 2))
}
```

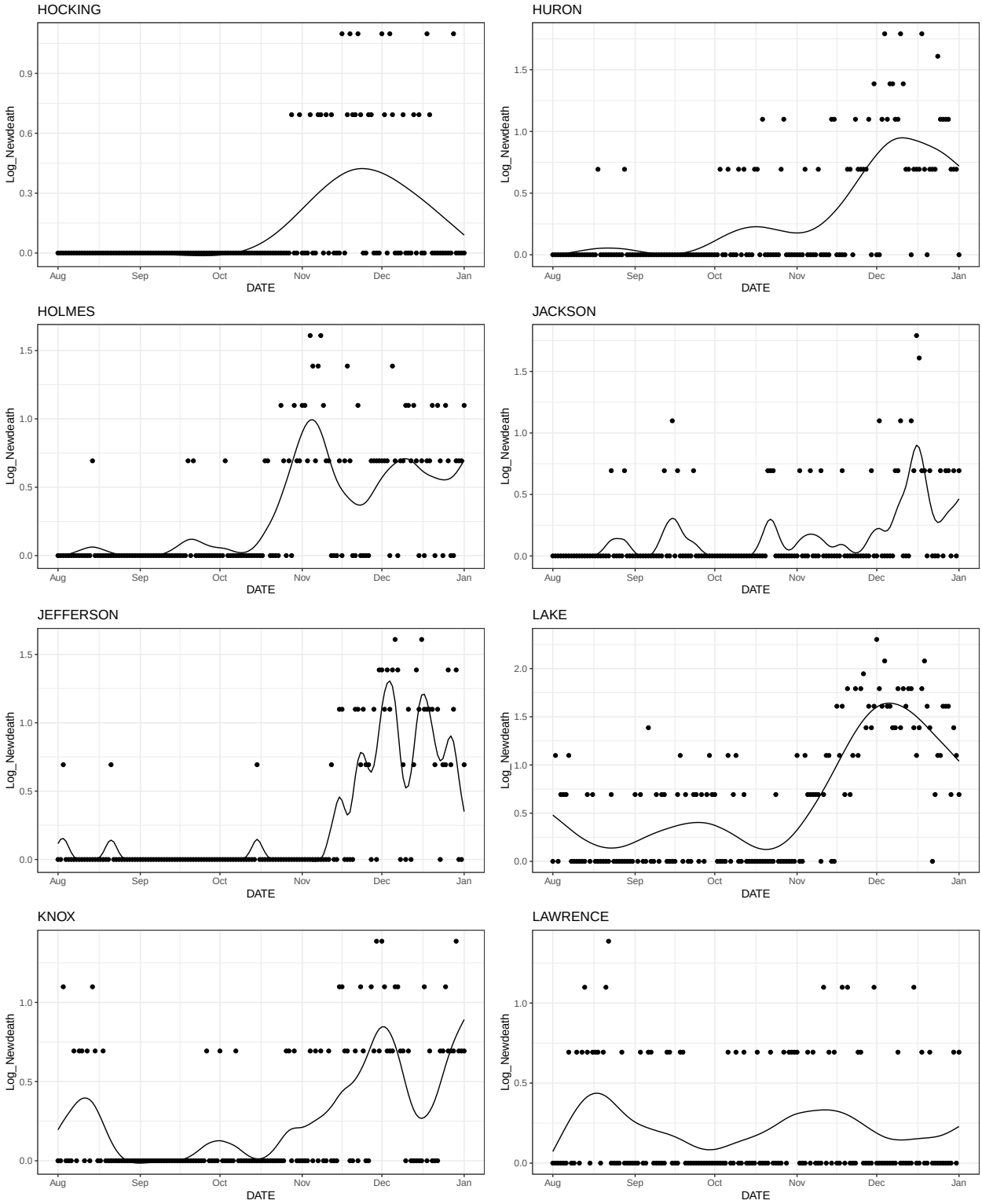


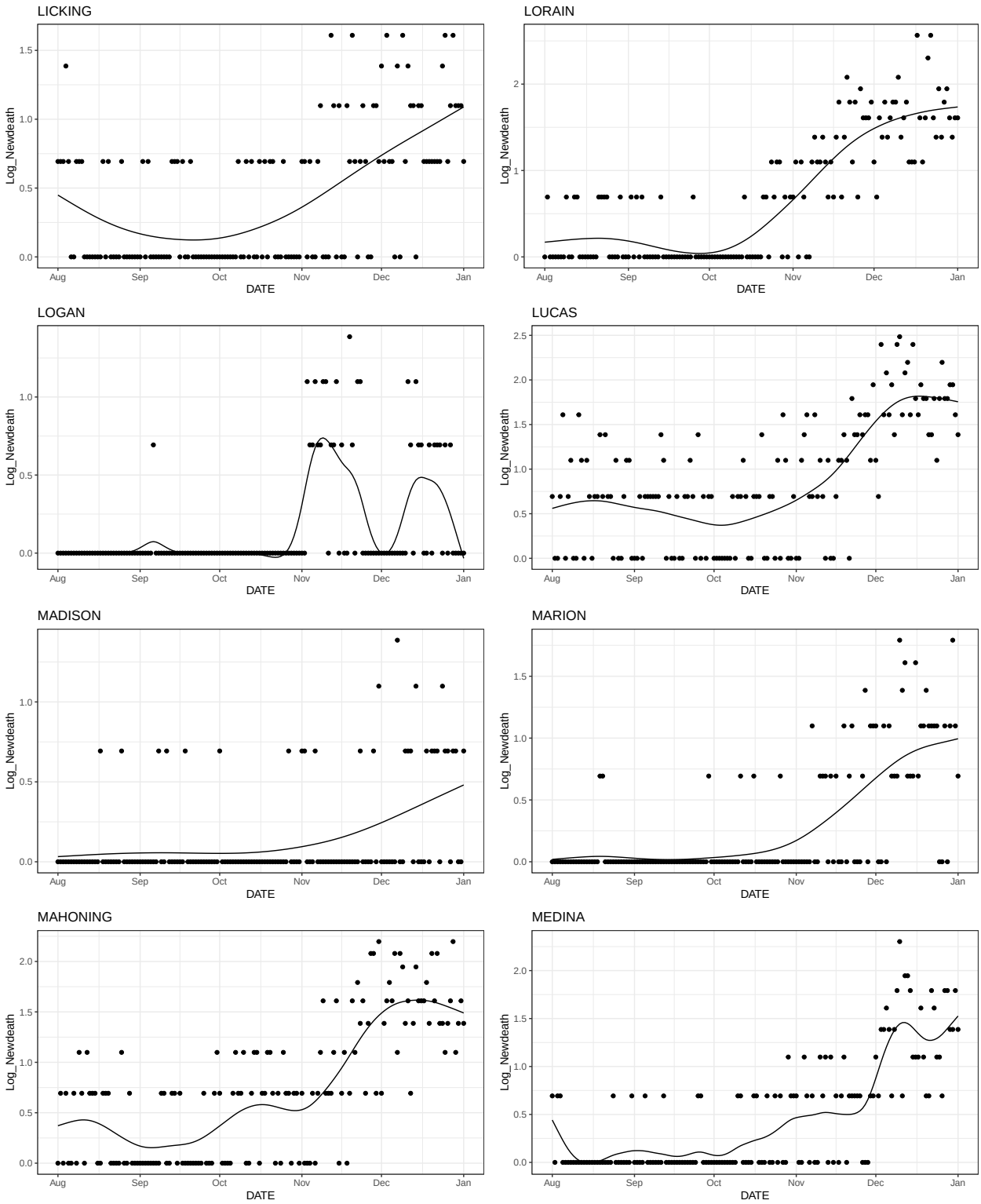

ATHENS
BELMONT
AUGLAIZE
BROWN
BUTLER
CHAMPAIGN
CARROLL
CLARK
Log_Newdeath
DATE
Aug
Sep
Oct
Nov
Dec
Jan

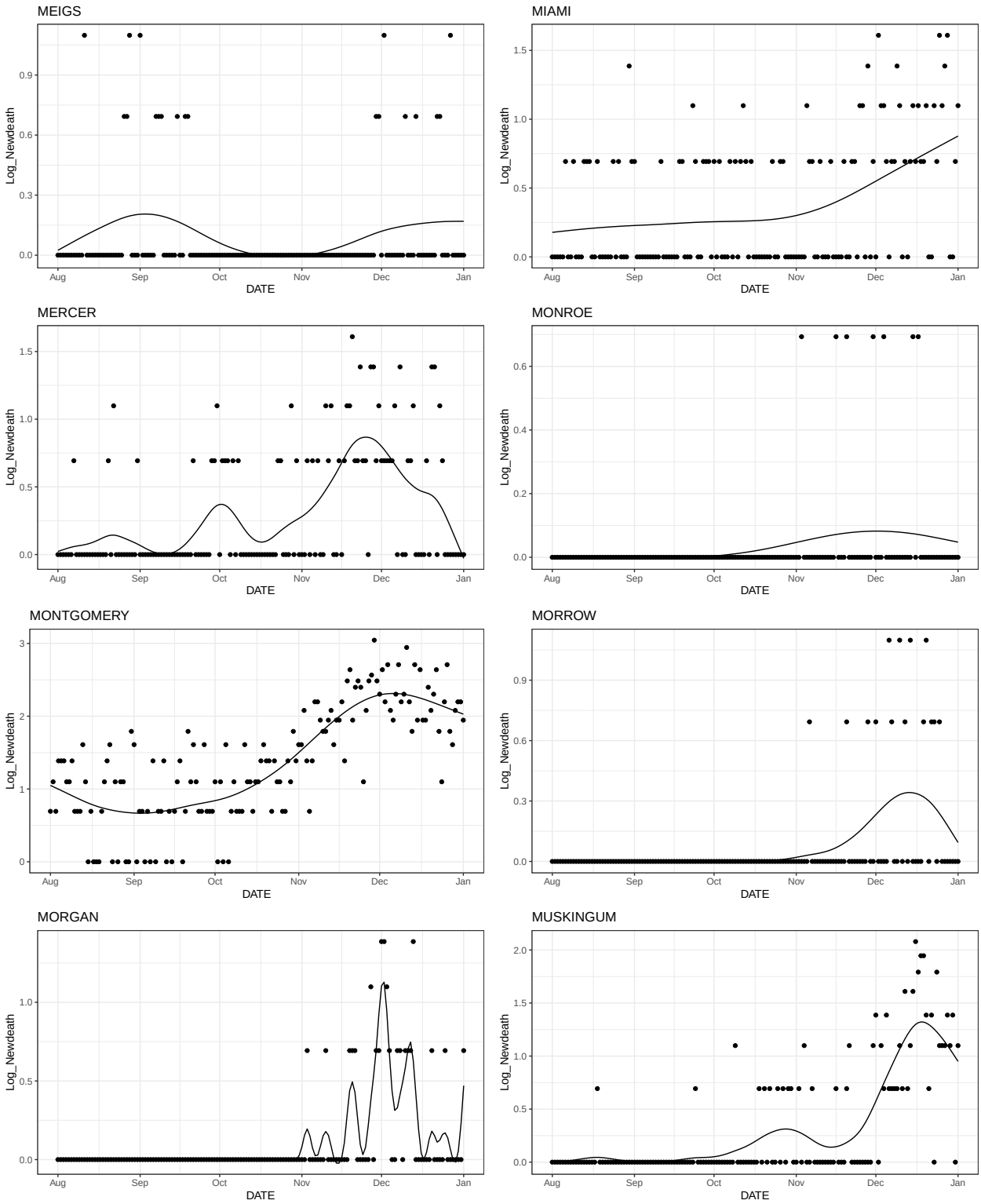

CLERMONT
COLUMBIANA
CLINTON
COSHOCTON
CRAWFORD
DARKE
CUYAHOGA
DEFIANCE
Log_Newdeath
DATE
Aug
Sep
Oct
Nov
Dec
Jan

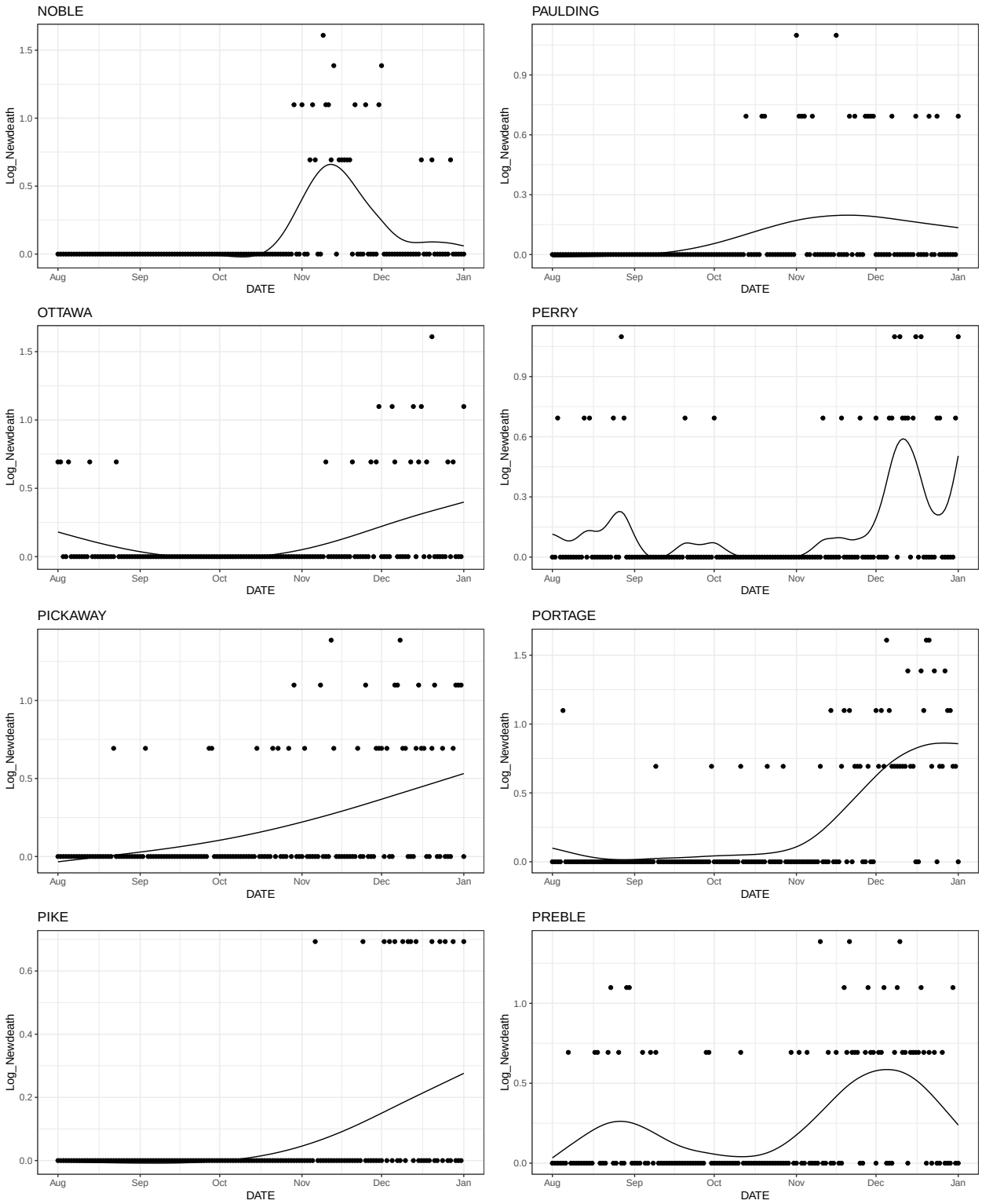

DELAWARE
FAIRFIELD
ERIE
FAYETTE
FRANKLIN
GALLIA
FULTON
GEAUGA
Log_Newdeath
DATE
Aug
Sep
Oct
Nov
Dec
Jan

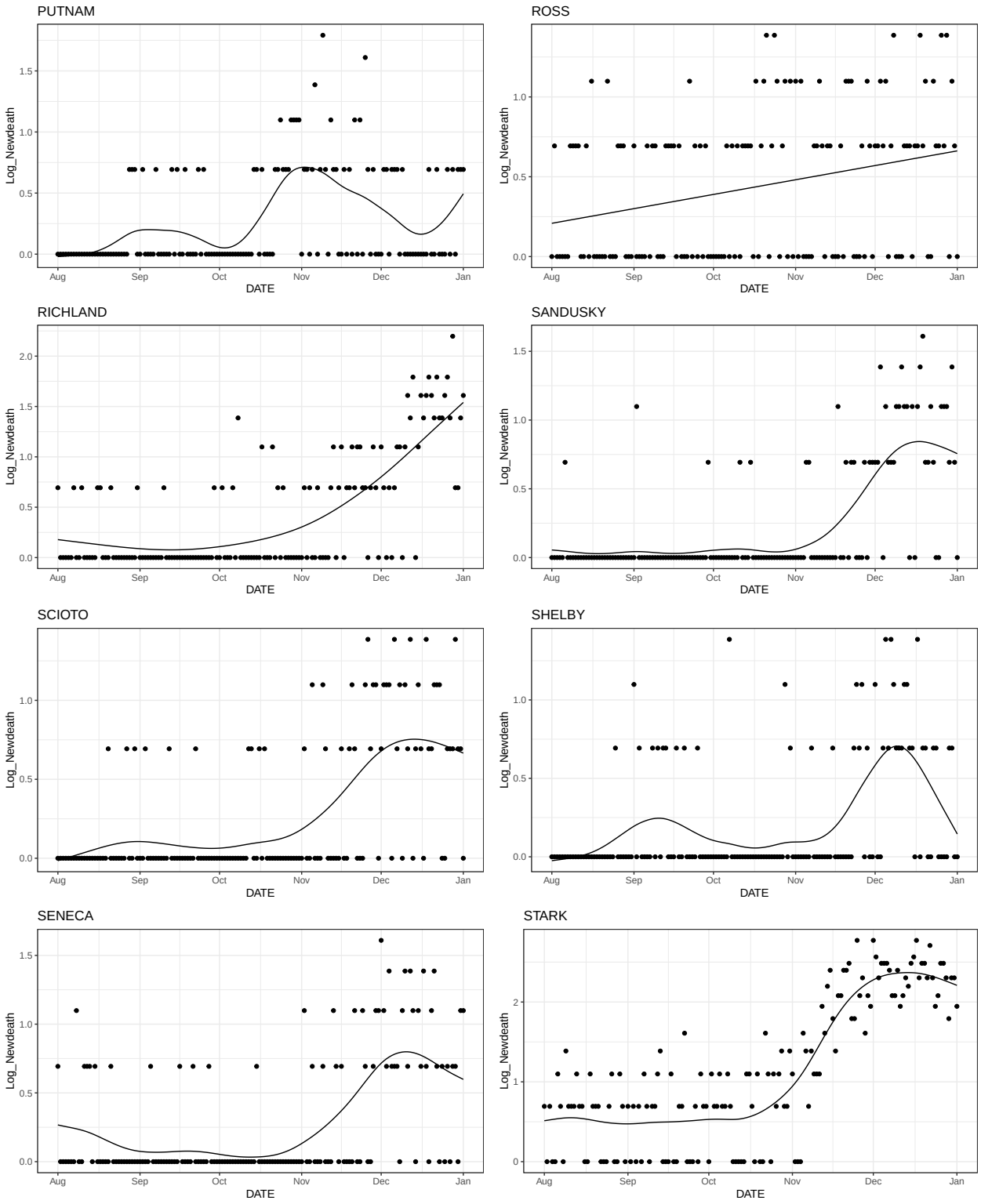

GREENE
HAMILTON
GUERNSEY
HANCOCK
HARDIN
HENRY
HARRISON
HIGHLAND
Log_Newdeath
DATE
Aug
Sep
Oct
Nov
Dec
Jan

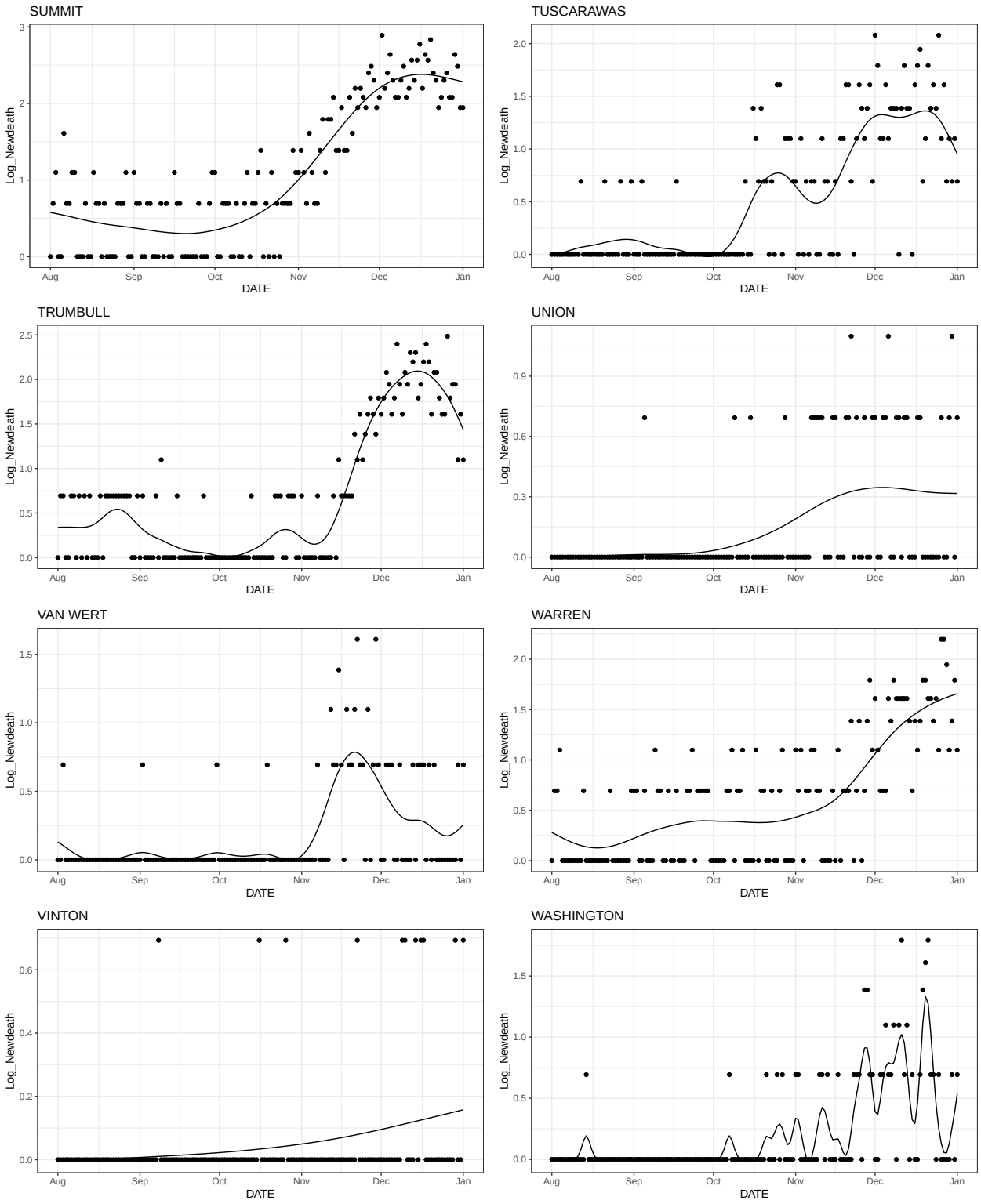

HOCKING
HURON
HOLMES
JACKSON
JEFFERSON
LAKE
KNOX
LAWRENCE
Log_Newdeath
DATE
Aug
Sep
Oct
Nov
Dec
Jan

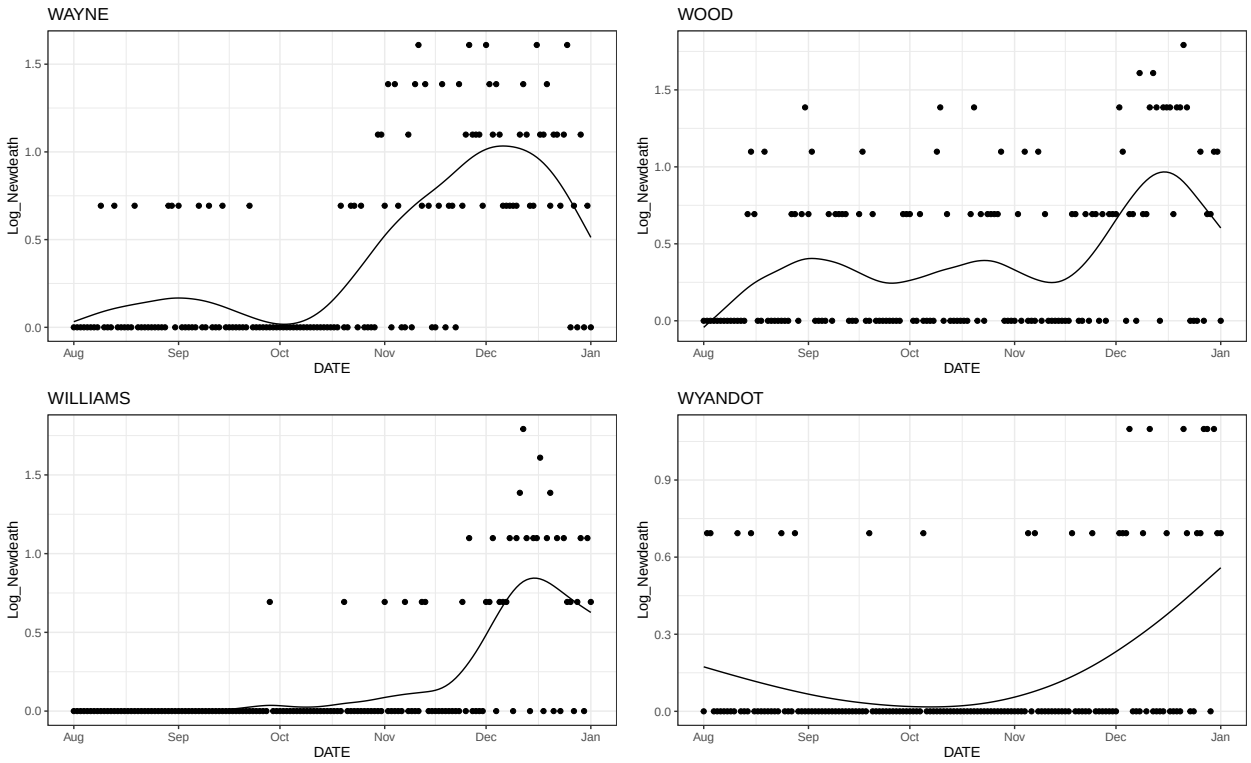

LICKING
LORAIN
LOGAN
LUCAS
MADISON
MARION
MAHONING
MEDINA
Log_Newdeath
DATE
Aug
Sep
Oct
Nov
Dec
Jan


MEIGS
MIAMI
MERCER
MONROE
MONTGOMERY
MORROW
MORGAN
MUSKINGUM
Log_Newdeath
DATE
Aug
Sep
Oct
Nov
Dec
Jan


NOBLE
PAULDING
OTTAWA
PERRY
PICKAWAY
PORTAGE
PIKE
PREBLE
Log_Newdeath
DATE
Aug
Sep
Oct
Nov
Dec
Jan


PUTNAM
ROSS
RICHLAND
SANDUSKY
SCIOTO
SHELBY
SENECA
STARK
Log_Newdeath
DATE
Aug
Sep
Oct
Nov
Dec
Jan


SUMMIT
TUSCARAWAS
TRUMBULL
UNION
VAN WERT
WARREN
VINTON
WASHINGTON
Log_Newdeath
DATE
Aug
Sep
Oct
Nov
Dec
Jan



Calculation of spline slope for the start of the semester

```
spline_slope_start<-c()
for(i in 1:88){
  county.data<-death%>%filter(COUNTY==county_name$COUNTY[i])
  spline<-smooth.spline(x = county.data$DATE,
                        y = county.data$Log_Newdeath)
  # Not sure whether we can calculate through this way
  slope.start<-spline$y[27]-spline$y[26]
  spline_slope_start<-c(spline_slope_start,slope.start)
}

spline_slop<-data.frame(
  COUNTY=county_name,
  spline_slop_start=round(spline_slope_start,4)
)
head(spline_slop,15)
```

```
##        COUNTY spline_slop_start
## 1       ADAMS           -0.0043
## 2       ALLEN           -0.0064
## 3     ASHLAND            0.0004
## 4   ASHTABULA            0.0016
## 5      ATHENS            0.0006
## 6    AUGLAIZE            0.0016
## 7     BELMONT           -0.0007
## 8       BROWN            0.0010
## 9      BUTLER            0.0076
## 10    CARROLL            0.0008
```

```
## 11  CHAMPAIGN           0.0028
## 12      CLARK          -0.0512
## 13   CLERMONT           0.0047
## 14    CLINTON           0.0023
## 15 COLUMBIANA           0.0196
```