BIOS 661 Project 1 Draft

Xinran Chu

9/22/2019

```
dat = read_tsv("https://raw.githubusercontent.com/biodatascience/datasci611/gh-pages/data/project1_2019/

dat = dat[ ,-c(2,3,4,5,8,9,10,11,12,13,14,15,16,17,18)]

dat$Date =  as.Date(dat$Date,'%m/%d/%Y')

dat.food = dat[ ,-2]

dat.food.before.2000 = dat.food %>%
  filter(substr(dat.food$Date, 1, 4) < 2000 & dat.food$`Food Pounds` > 0)


length(which(!is.na(dat.food.before.2000$`Food Pounds`)))
```

```
## [1] 47
```

```
# which(!is.na(dat.food.before.2000$`Food Pounds`))

dat.food.2000to2009 = dat.food %>%
  filter(substr(dat.food$Date, 1, 4) < 2010 & substr(dat.food$Date, 1, 4) >= 2000 &   dat.food$`Food Pou
#View(dat.food.2000to2009)

length(which(!is.na(dat.food.2000to2009$`Food Pounds`)))
```

```
## [1] 8822
```

```
#which(!is.na(dat.food.2000to2009$`Food Pounds`))

dat.food.after2009 = dat.food %>%
  filter(substr(dat.food$Date, 1, 4) >= 2010 &   dat.food$`Food Pounds` > 0)
#View(dat.food.after2009)

length(which(!is.na(dat.food.after2009$`Food Pounds`)))
```

```
## [1] 40448
```

```
#which(!is.na(dat.food.after2009$`Food Pounds`))
```

```
summary(dat.food.before.2000$`Food Pounds`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.00   15.00   19.00   25.23   30.00   60.00
```

```
summary(dat.food.2000to2009$`Food Pounds`)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00   10.00   15.00   18.15   25.00 1700.00
```

```
summary(dat.food.after2009$`Food Pounds`)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
```

```
##      1.0     15.0     20.0     38.3     30.0 450121.0
```

```
# total lbs of food provided inn each period
sum(dat.food.before.2000$`Food Pounds`) # 1186 lbs of food provided before 2000
```

```
## [1] 1186
```

```
sum(dat.food.2000to2009$`Food Pounds`) # 160116 lbs of food provided between 2000 and 2009
```

```
## [1] 160116
```

```
sum(dat.food.after2009$`Food Pounds`) # 1606101 lbs of food provided on and after 2010
```

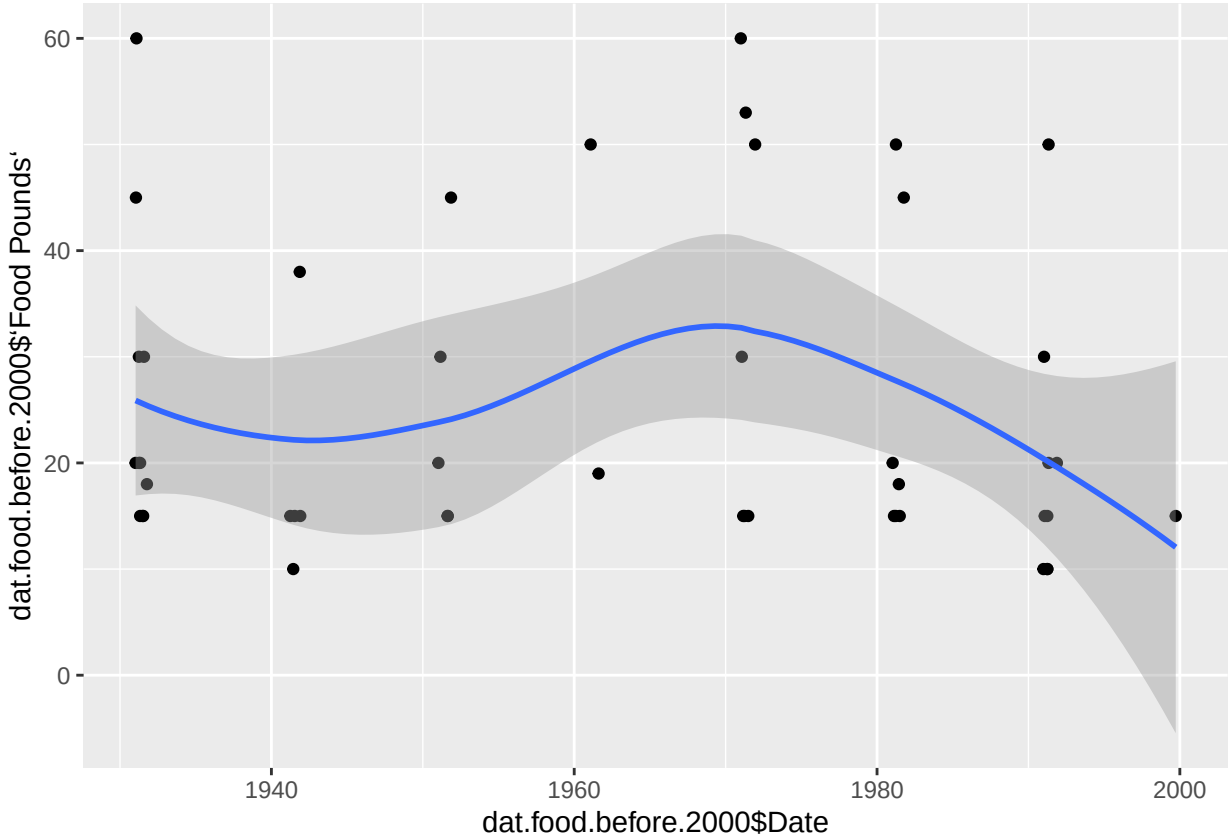
```
## [1] 1549468
```

```
# Observe a large increase in the amount of food provided over the years.
```

Plot the trend for each period

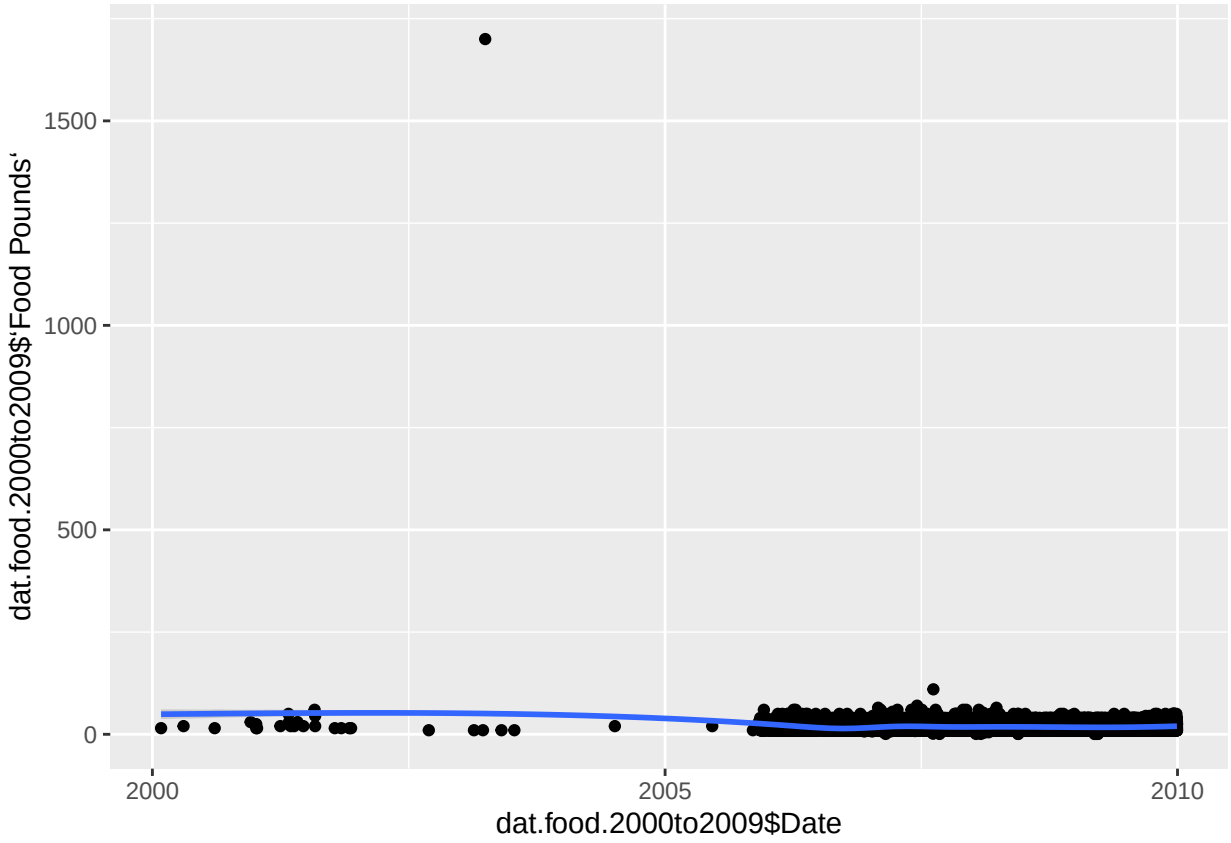
```
ggplot(dat.food.before.2000,aes(x=dat.food.before.2000$Date, y=dat.food.before.2000$`Food Pounds`)) +
  geom_point()+
  geom_smooth()
```

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



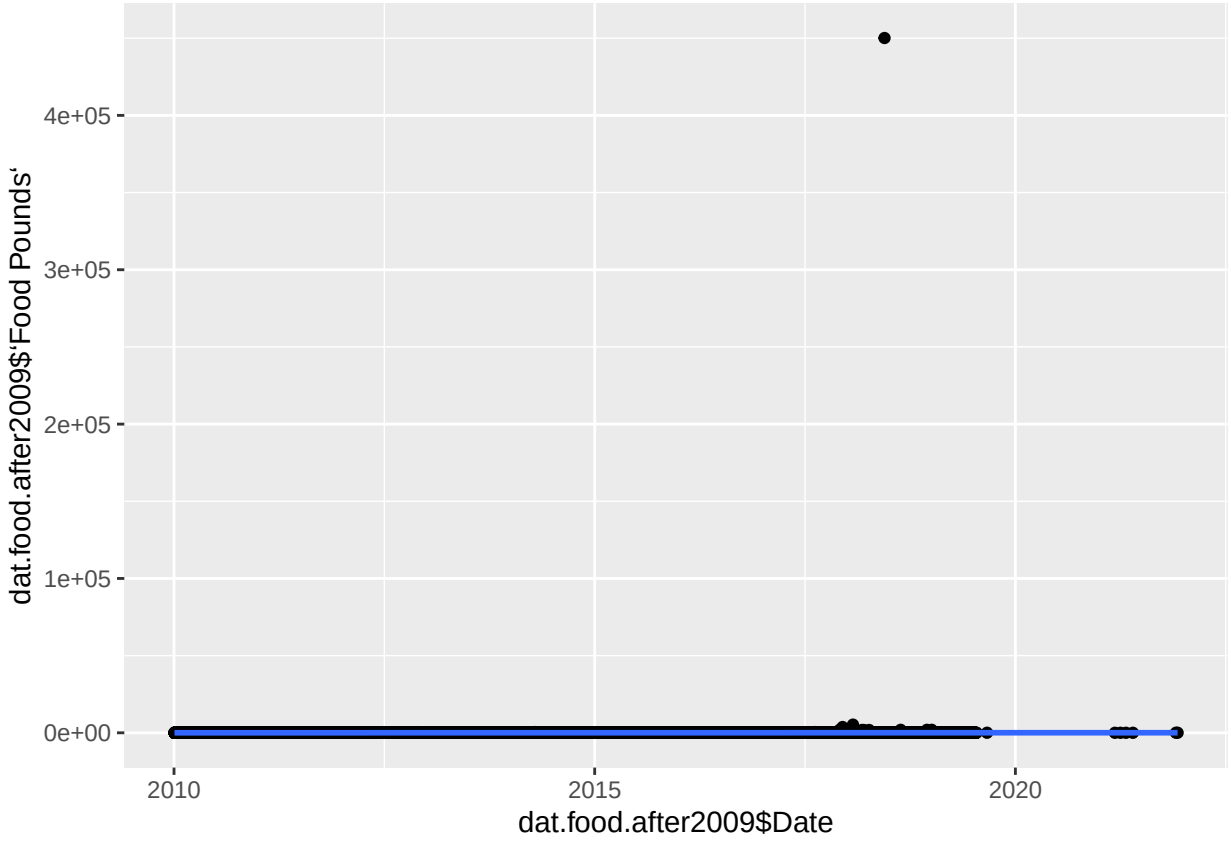

```
ggplot(dat.food.2000to2009,aes(x=dat.food.2000to2009$Date, y=dat.food.2000to2009$`Food Pounds`)) +
  geom_point() +
  geom_smooth()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
ggplot(dat.food.after2009,aes(x=dat.food.after2009$Date, y=dat.food.after2009$`Food Pounds`)) +
  geom_point() +
  geom_smooth()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```



Number of People for Which Food Was Provided

```
dat.people = dat[ ,-3]

dat.people.before.2000 = dat.people %>%
  filter(substr(dat.people$Date, 1, 4) < 2000 & dat.people$`Food Provided for` > 0)

dat.people.2000to2009 = dat.people %>%
  filter(substr(dat.people$Date, 1, 4) < 2010 & substr(dat.people$Date, 1, 4) >= 2000  &   dat.people$`

dat.people.after2009 = dat.people %>%
  filter(substr(dat.people$Date, 1, 4) >= 2010 & dat.people$`Food Provided for` > 0)
```

```
summary(dat.people.before.2000)
```

```
##       Date            Food Provided for
##  Min.   :1931-01-12   Min.   : 1.000
##  1st Qu.:1999-01-08   1st Qu.: 1.000
##  Median :1999-05-25   Median : 1.000
##  Mean   :1995-04-04   Mean   : 2.035
##  3rd Qu.:1999-09-12   3rd Qu.: 3.000
##  Max.   :1999-12-21   Max.   :23.000
```

```
summary(dat.people.2000to2009)
```

```
##       Date            Food Provided for
```

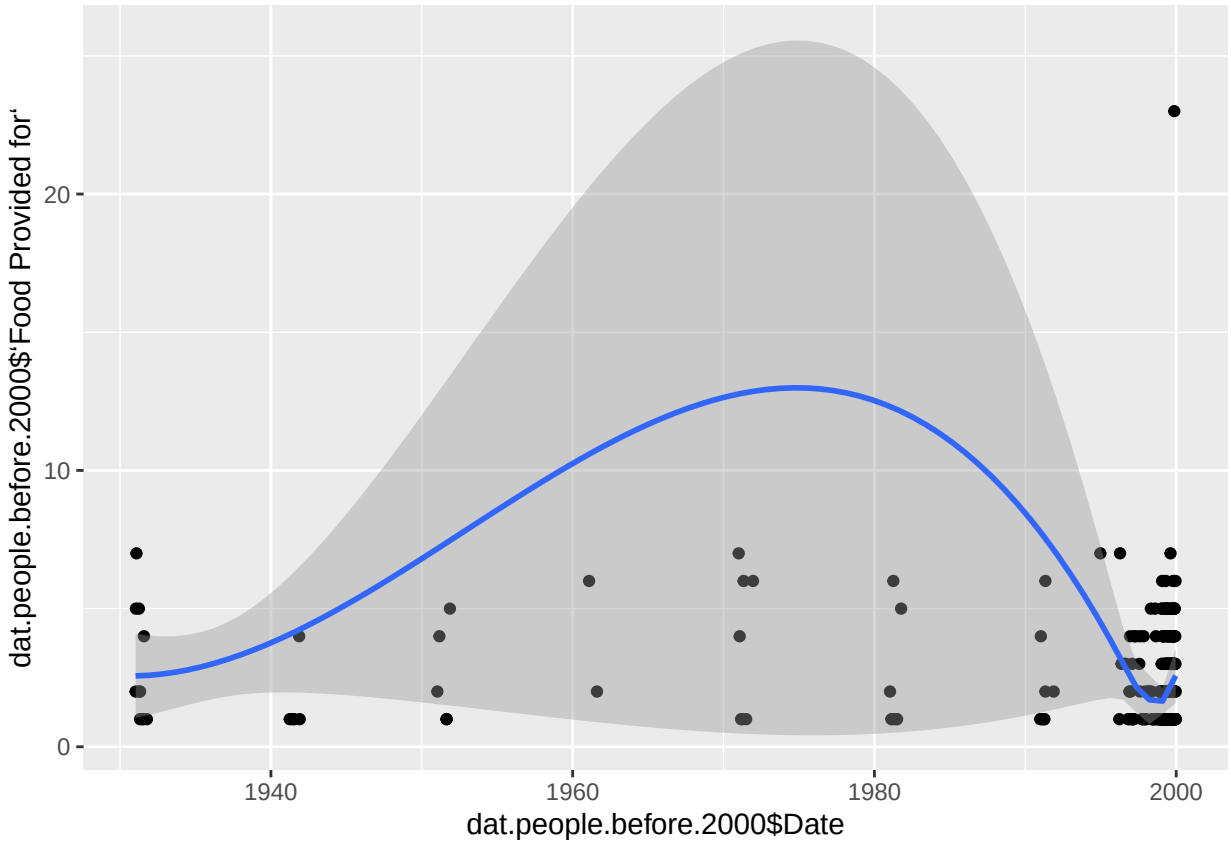
```
##  Min.   :2000-01-04   Min.   :  1.000
##  1st Qu.:2003-05-19   1st Qu.:  1.000
##  Median :2006-08-15   Median :  2.000
##  Mean   :2006-01-29   Mean   :  2.562
##  3rd Qu.:2008-09-16   3rd Qu.:  4.000
##  Max.   :2009-12-30   Max.   :750.000
```

```
summary(dat.people.after2009)
```

```
##       Date            Food Provided for
##  Min.   :2010-01-03   Min.   :   1.000
##  1st Qu.:2012-09-26   1st Qu.:   1.000
##  Median :2015-02-11   Median :   2.000
##  Mean   :2015-01-07   Mean   :   2.863
##  3rd Qu.:2017-05-17   3rd Qu.:   4.000
##  Max.   :2021-12-05   Max.   :1151.000
```

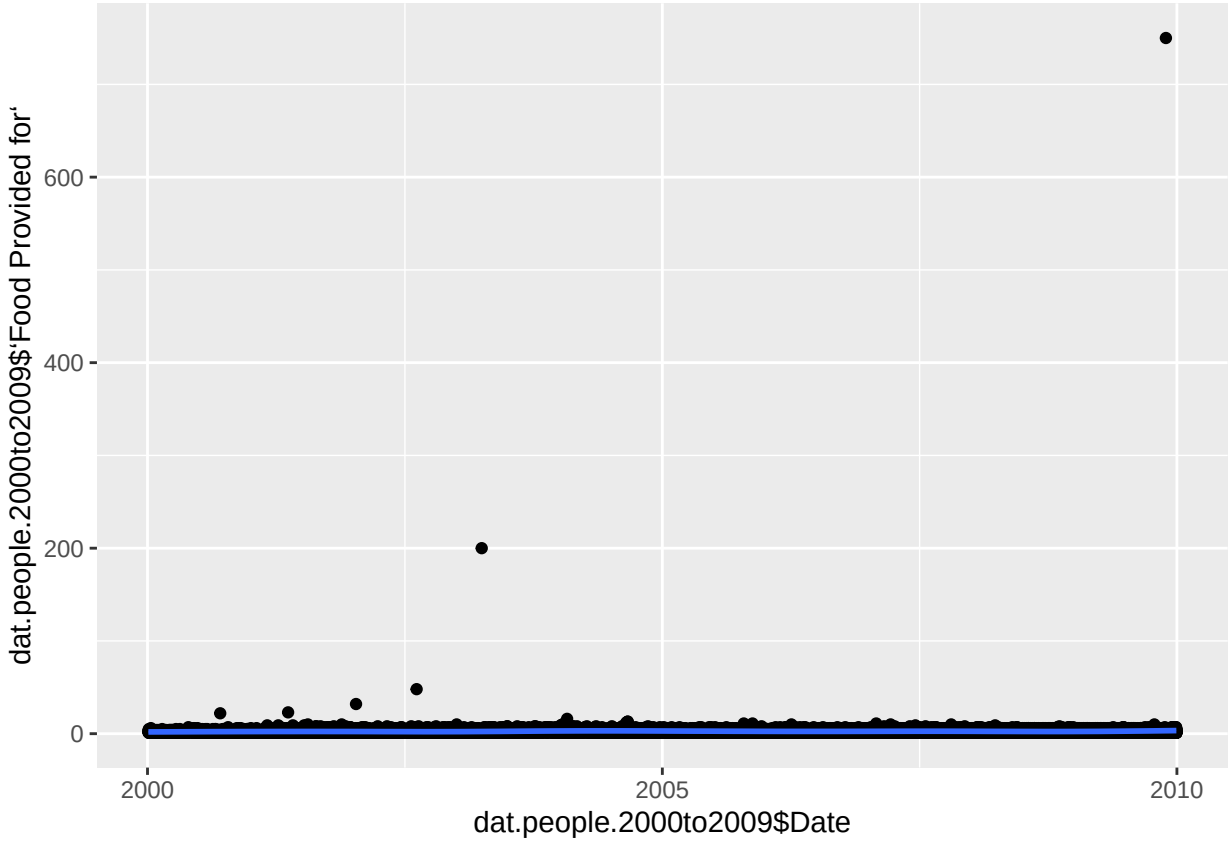
```
ggplot(dat.people.before.2000,aes(x=dat.people.before.2000$Date, y=dat.people.before.2000$`Food Provide
  geom_point()+
  geom_smooth()
```

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



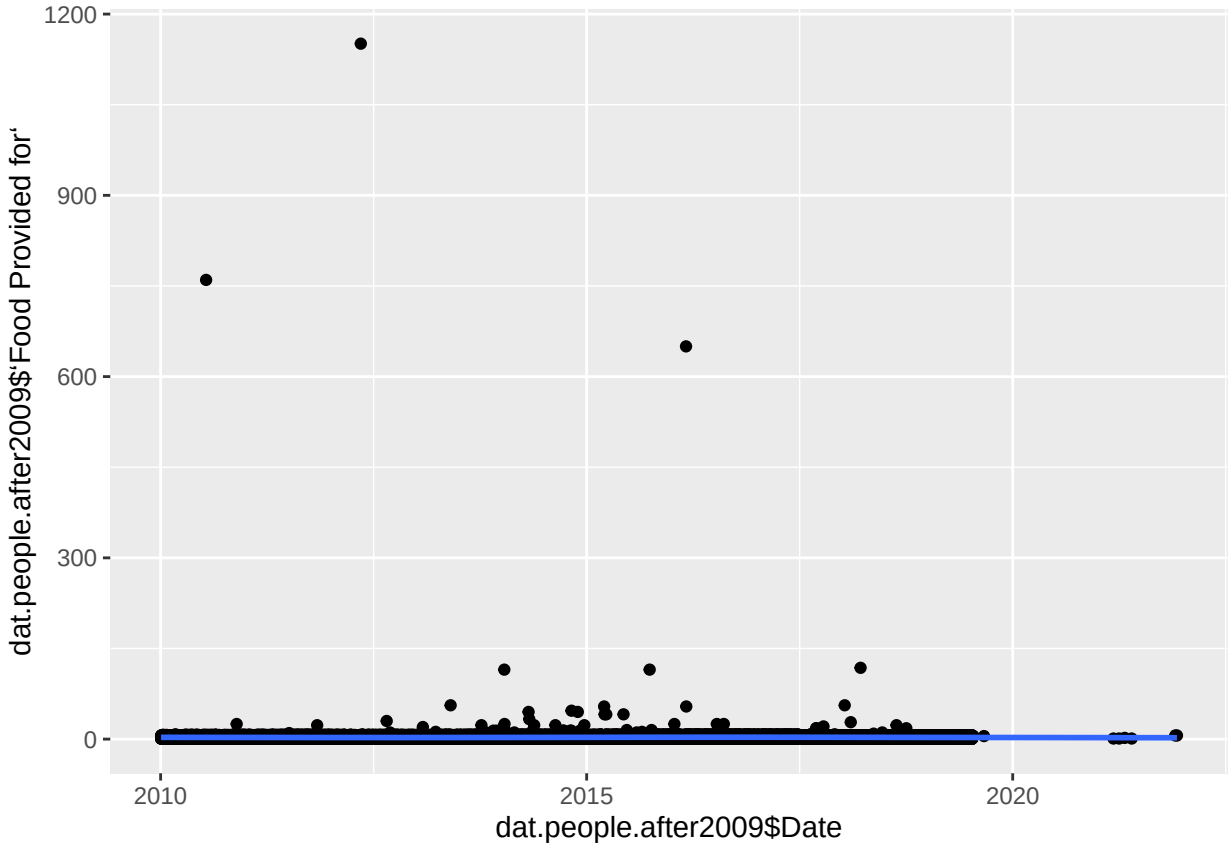

```
ggplot(dat.people.2000to2009,aes(x=dat.people.2000to2009$Date, y=dat.people.2000to2009$`Food Provided f
  geom_point() +
  geom_smooth()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```




```
ggplot(dat.people.after2009,aes(x=dat.people.after2009$Date, y=dat.people.after2009$`Food Provided for`)
  geom_point() +
  geom_smooth()
```

```
## `geom_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```


1200
900
600
300
0
dat.people.after2009$`Food Provided for`
2010
2015
2020
dat.people.after2009$Date