

Nyc flights

```
library ( nycflights13 )
```

```
library ( tidyverse )
```

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

```
## v ggplot2 3.3.0      v purrr   0.3.4
## v tibble  3.0.1      v dplyr   0.8.5
## v tidyr   1.0.3      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(ggplot2)
```

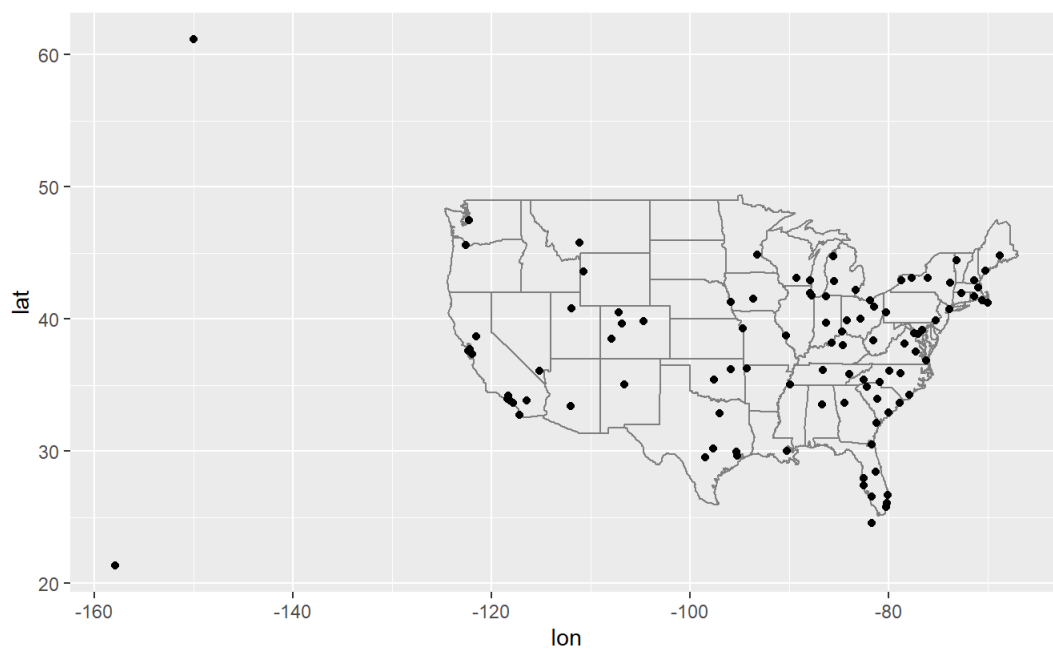
```
library(maps)
```

```
##
## Attaching package: 'maps'
```

```
## The following object is masked from 'package:purrr':
##
##      map
```

```
library(dplyr)
```

```
## question -1
flights<-nycflights13::flights
airports<-nycflights13::airports
planes<-nycflights13::planes
weather<-nycflights13::weather
avgdelay<-group_by(flights,dest)%>%
  summarise(avg_delay=mean(arr_delay,na.rm=TRUE))
airports<-rename(airports,"dest"="faa")
join1<-airports%>%semi_join(avgdelay,by='dest')
airports%>%semi_join(avgdelay,by='dest')%>%ggplot(aes(lon,lat,colors="avg_delay"))+borders("state")+geom_poi
nt()+coord_quickmap()
```



##2nd question

```
loc<-airports%>%select(dest,lat,lon)
flights_loc<-flights%>%left_join(loc,by=c('origin'='dest'))
flights_loc<-rename(flights_loc,"origin_lat"="lat","origin_lon"="lon")
flights_loc<-flights_loc%>%left_join(loc,by="dest")
flights_loc<-rename(flights_loc,"dest_lat"="lat","dest_lon"="lon")
```

##3rd question

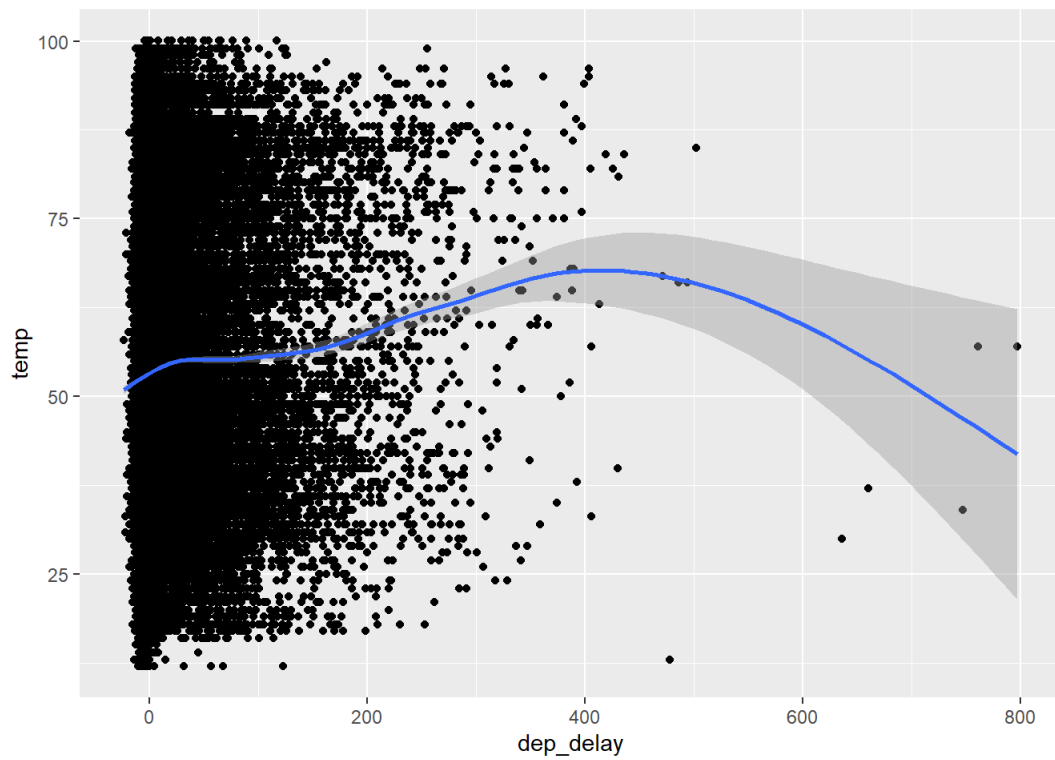
```
planes_age<-planes%>%select(tailnum,year)
planes_age<-arrange(planes_age,-desc(year))
flights_delay<-group_by(flights,tailnum)%>%
  summarise(delay=mean(arr_delay,na.rm=TRUE))
flights_age<-flights_delay%>%inner_join(planes_age,by="tailnum")
flights_age%>%drop_na()
```

```
## # A tibble: 3,246 x 3
##   tailnum  delay  year
##   <chr>    <dbl> <int>
## 1 N10156  12.7    2004
## 2 N102UW   2.94    1998
## 3 N103US  -6.93    1999
## 4 N104UW   1.80    1999
## 5 N10575  20.7    2002
## 6 N105UW  -0.267   1999
## 7 N107US  -5.73    1999
## 8 N108UW  -1.25    1999
## 9 N109UW  -2.52    1999
## 10 N110UW   2.8     1999
## # ... with 3,236 more rows
```

##4th question

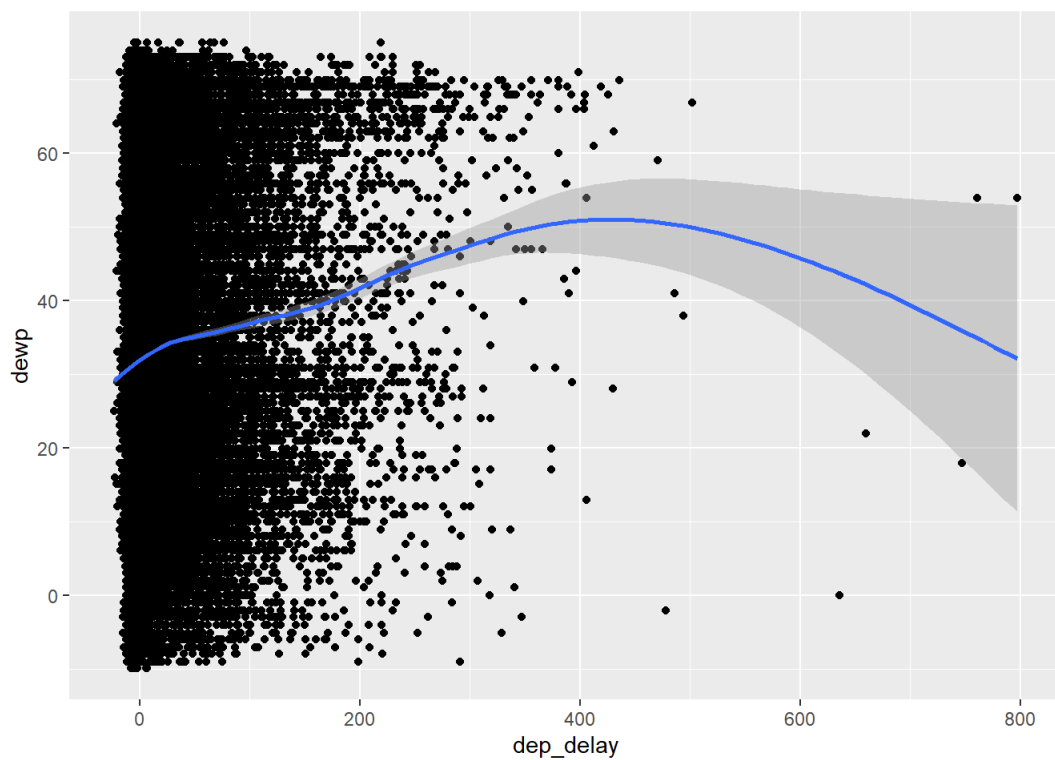
```
by_day <-group_by ( flights , year , month , day )
flights_weather<-merge(weather,flights,by=c("origin",'year','month','day','hour'))
flights_weather<-flights_weather%>%drop_na()
ggplot(data=flights_weather,aes(x=dep_delay,y=temp))+geom_point()+geom_smooth()
```

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



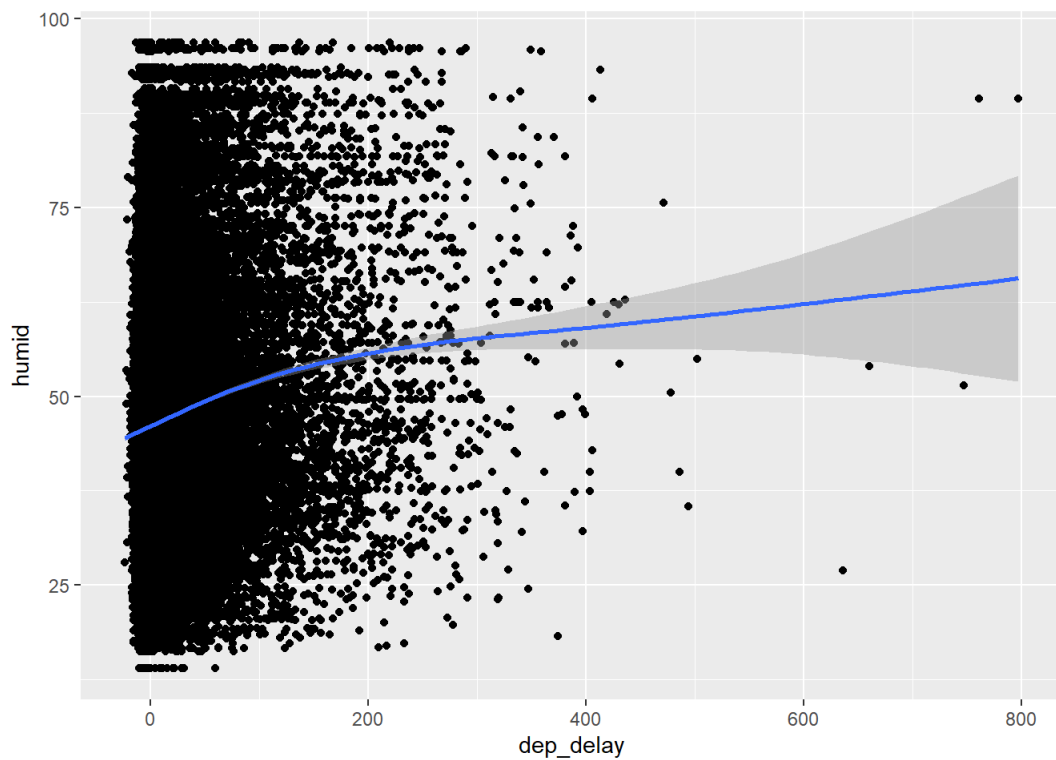
```
ggplot(data=flights_weather, aes(x=dep_delay, y=temp)) + geom_point() + geom_smooth()
```

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



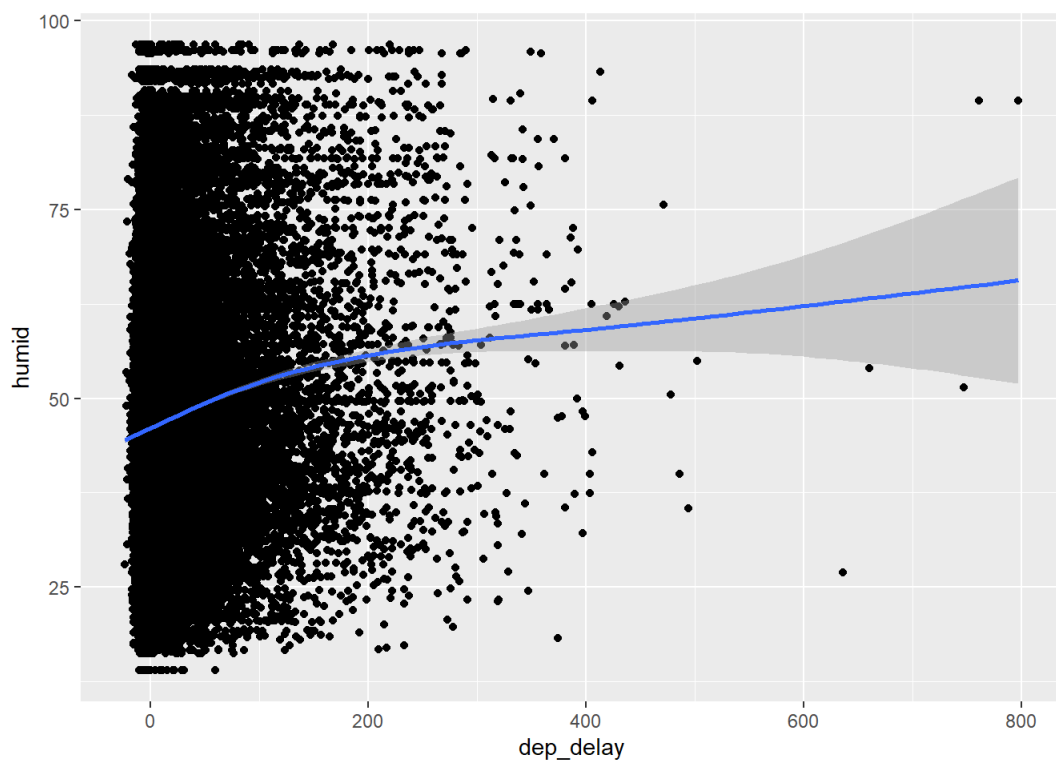
```
ggplot(data=flights_weather, aes(x=dep_delay, y=dewp)) + geom_point() + geom_smooth()
```

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



```
ggplot(data=flights_weather,aes(x=dep_delay,y=humid))+geom_point()+geom_smooth()
```

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



```
##5th question
june13<-filter(flights,year==2013,month==6,day==13)
flights_weather<-merge(june13,weather,by=c('year','month','day','hour'))
ggplot(june13,aes(x=hour,y=dep_delay))+geom_point()+geom_smooth()
```

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

```
## Warning: Removed 95 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 95 rows containing missing values (geom_point).
```

