Plotting groups with "facets"

Facets are different faces of the data, corresponding to "group_by" subsets.

Can plot each facet separately in ggplot.
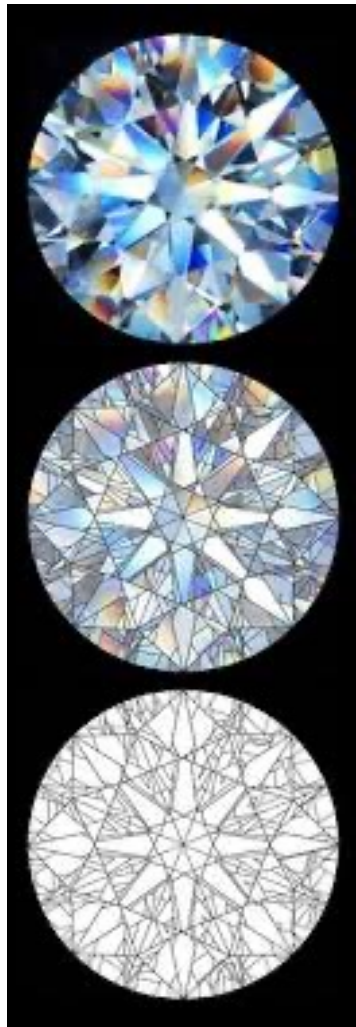
commands: facet_grid (2d) and facet_wrap (1d)

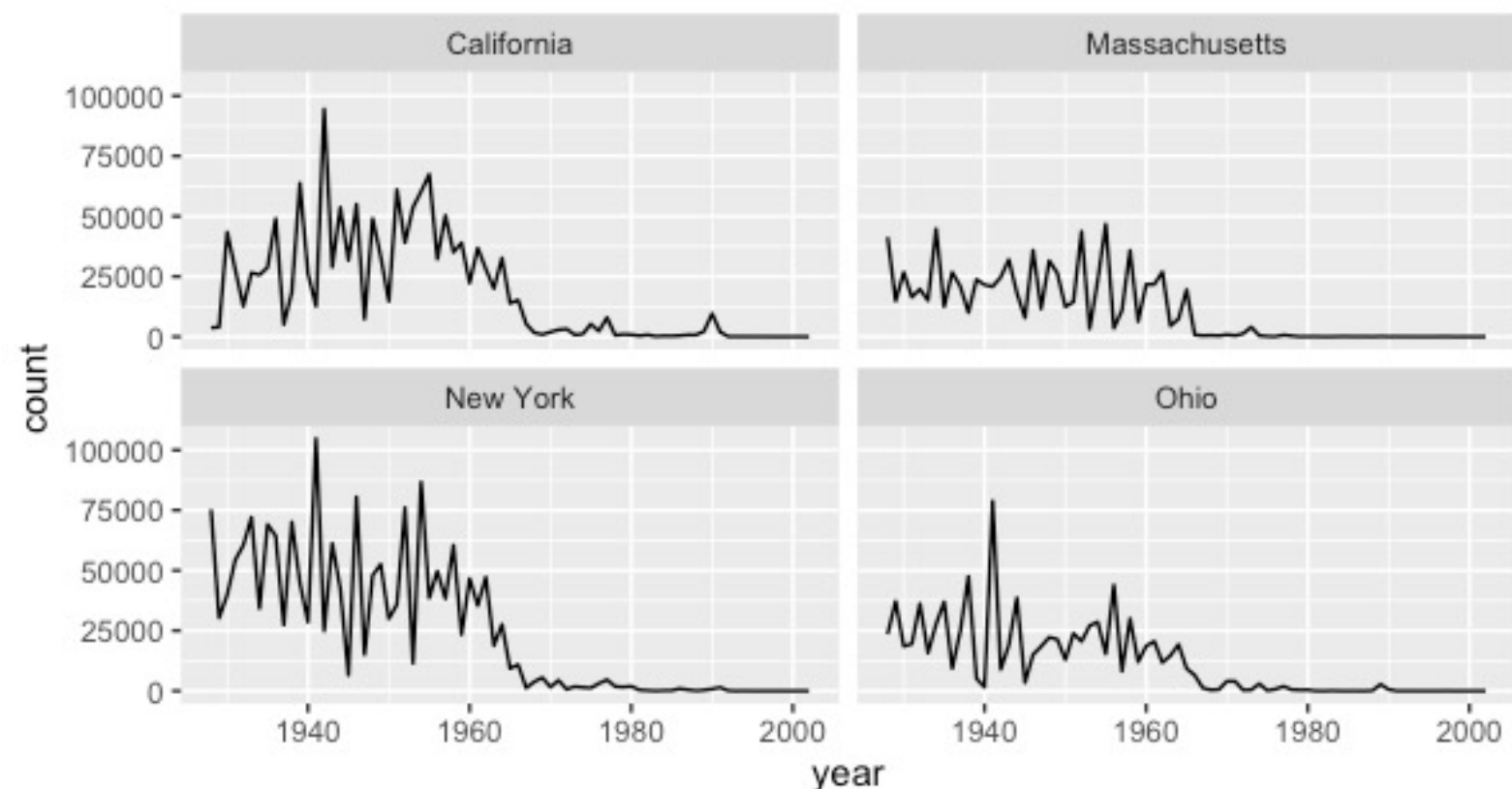for facet_grid, specify a formula,
   `vertical-var ~ horizontal-var`

for facet_wrap with one variable, the formula is just ~ `variable`

Plotting groups with "facets"

```
meas.small=filter(diseases,disease=="Measles",
  state %in% c("California", "Massachusetts",
  "New York","Ohio"))

ggplot(meas.small,aes(x=year,y=count))+
geom_line()+facet_wrap(~ state)
```
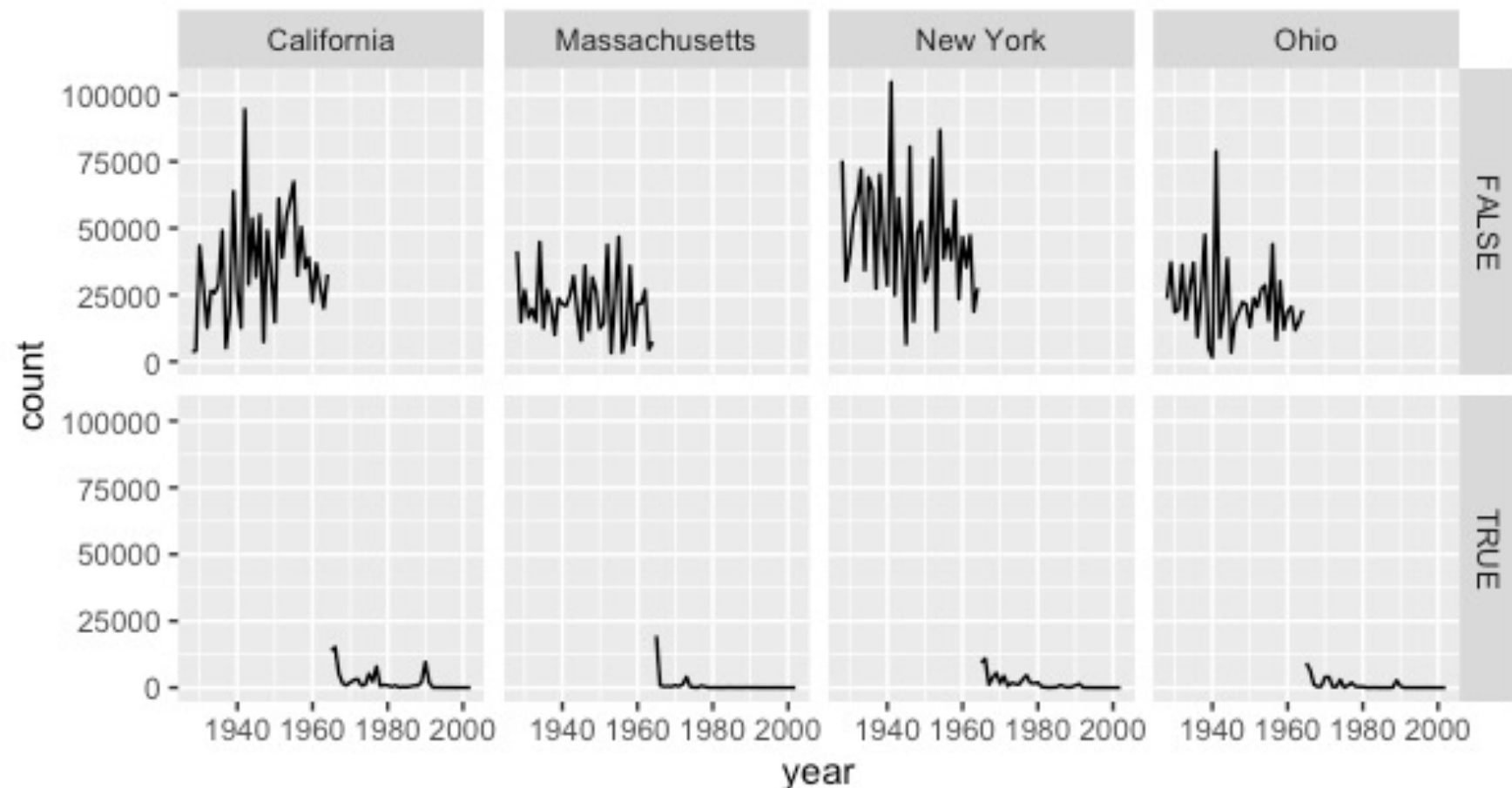
Plotting groups with "facets"

```
meas.small = mutate(meas.small,
                postvaccine=(year>1964))
ggplot(meas.small,aes(x=year,y=count))+
        geom_line()+facet_grid(postvaccine~state)
```

Diabetes data (background)

Pima Indian diabetes data set (UCIML; Kaggle)

A data frame with patient data from NIDDK.

```
pregnant:     Number of (prior) pregnancies
glucose:      Plasma glucose concentration at 2 hours
pressure:     Diastolic blood pressure (mm Hg)
triceps:      Triceps skin fold thickness (mm)
insulin:      2 hour serum insulin (muU/ml)
mass:         Body mass index, or BMI
pedigree:     Indication of family history of diabetes
age:          Patient age in years
diabetes:     Logical variable indicating a diabetes diagnosis
```

Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., & Johannes, R.S. (1988). Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In Proceedings of the Symposium on Computer Applications and Medical Care (pp. 261--265). IEEE Computer Society Press.

diabetes data set

```
library("mlbench")

data("PimaIndiansDiabetes2")

diab <- PimaIndiansDiabetes2
```

Can also load from
diabetes.rds on Schedule

PimaIndiansDiabetes {mlbench}

Pima Indians Diabetes Database

Description

A data frame with 768 observations on 9 variables.

Usage

```
data(PimaIndiansDiabetes)
data(PimaIndiansDiabetes2)
```

Format

pregnant Number of times pregnant
  glucose Plasma glucose concentration (glucose tolerance test)
pressure Diastolic blood pressure (mm Hg)
   triceps Triceps skin fold thickness (mm)
    insulin 2-Hour serum insulin (mu U/ml)
     mass Body mass index (weight in kg/(height in m)\^2)
pedigree Diabetes pedigree function
      age Age (years)
diabetes Class variable (test for diabetes)

Facets on diabetes data

Use `mutate` and `ifelse` (or `switch`) statements to create a new copy of `diab` called `diabp` that includes a categorical variable, `pregcat`, that maps the number of pregnancies into three possible groups:

pregnancies	category
0	"none"
1-4	"few"
5-8	"many"
>8	"lots"

Use `facet_wrap` to plot histograms of BMI ("mass") by category.

If you have extra time, make the bin width 5, and see if you can figure out how to display the facets in the order listed above.