

ggplot

<https://mnielsen.shinyapps.io/ggplot-slides>

CMS Public Data

The data used in these slides is public data located on the CMS page here. I've arranged the data into a more usable format for this presentation which can be downloaded by clicking the button below.

age	chronicCondition	state	prevalence	gender	year
All	Alzheimer's Disease/Dementia	National	8.09	Male	2007
All	Alzheimer's Disease/Dementia	Alabama	7.85	Male	2007
All	Alzheimer's Disease/Dementia	Alaska	5.86	Male	2007
All	Alzheimer's Disease/Dementia	Arizona	5.89	Male	2007
All	Alzheimer's Disease/Dementia	Arkansas	8.32	Male	2007
All	Alzheimer's Disease/Dementia	California	7.67	Male	2007

To read the data into R run the following code:

```
setwd("C:\\path\\to\\folder")
# setwd("C:/path/to/folder")
ccdat <- read.csv("chronic-conditions.csv")
```

ggplot2 Basics

```
# If you haven't already install the package and load the library.
install.packages("ggplot2"); library(ggplot2)
```

ggplot()

```
# Add layers to ggplot() with '+'
ggplot([insert data here], aes(x=[x-var], y=[y-var], color=[color-var], [...])) + geom_[graph type]
```

Layers can include combinations of “geoms”, “stats”, “scales”, “faceting”, “themes”, etc.

qplot()

```
# Build a plot using the qplot() function
qplot(x=[x-var], y=[y-var], color=[color-var], [...], data=[insert data here], geom = "[graph type]")
```

“Some introductions to ggplot2 make use of a function called `qplot()`, which is intended as a convenient interface for making graphs. It does require a little less typing than using `ggplot()` plus a `geom`, but I’ve found it a bit confusing to use because it has a slightly different way of specifying certain graphing parameters. I think it’s simpler and easier to just use `ggplot()`.” -Winston Chang (R Graphics Cookbook, 2012)

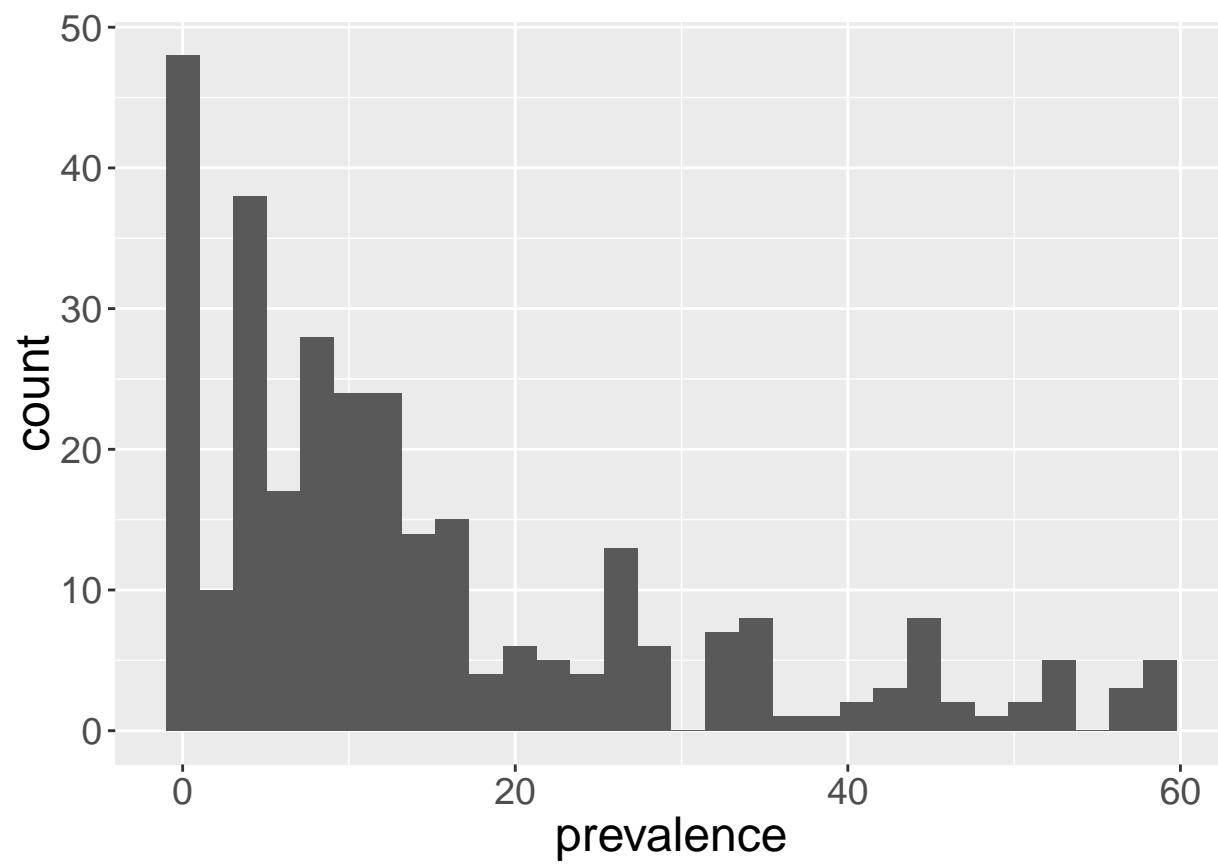
Geoms (One Variable)

```
library(ggplot2)
ccdat <- read.csv("../chronic-conditions.csv")

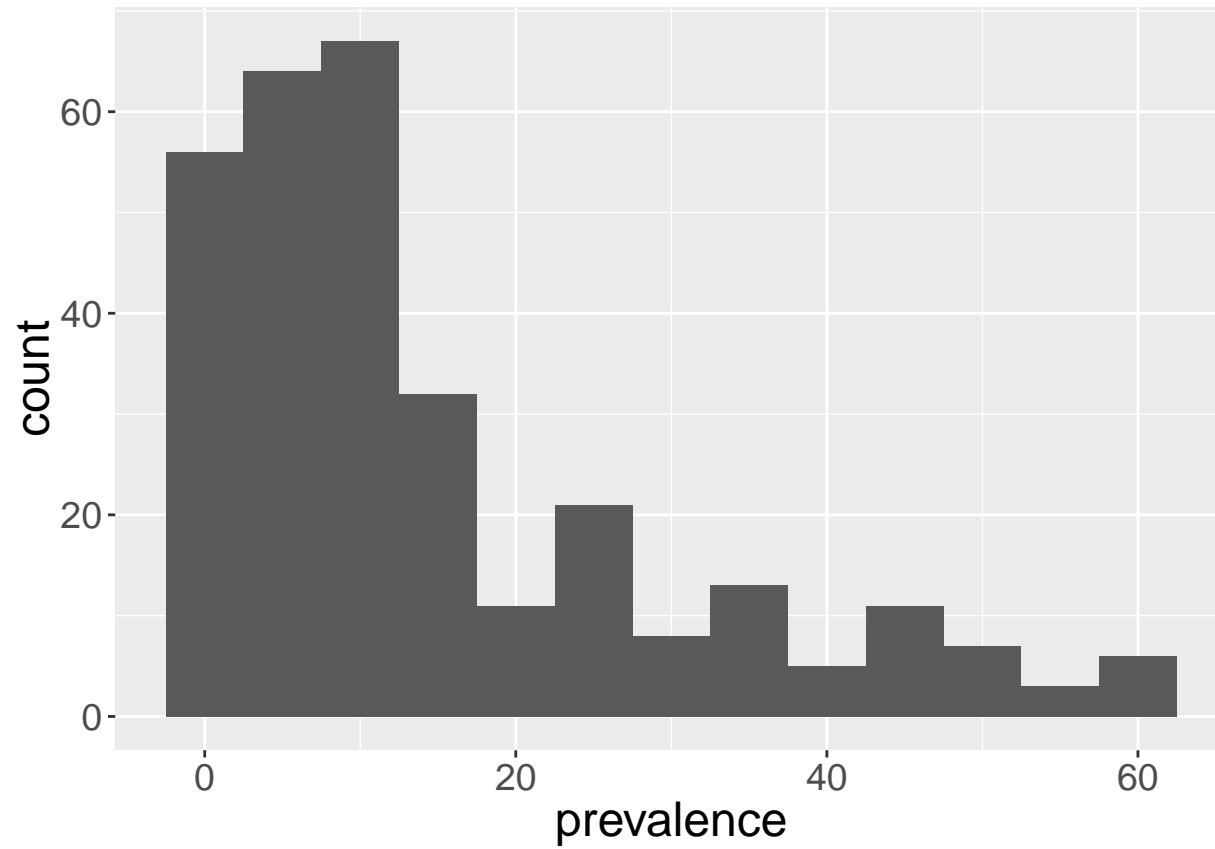
new.theme <- theme_update(axis.text = element_text(size=14),
                          axis.title=element_text(size=18))
trim <- function (x) gsub("^\\s+|\\s+$", "", x)
ccdat$state <- trim(ccdat$state)
ccdat$year <- as.numeric(ccdat$year)

alldat <- ccdat[ccdat$age=="All" & ccdat$state=="National",]

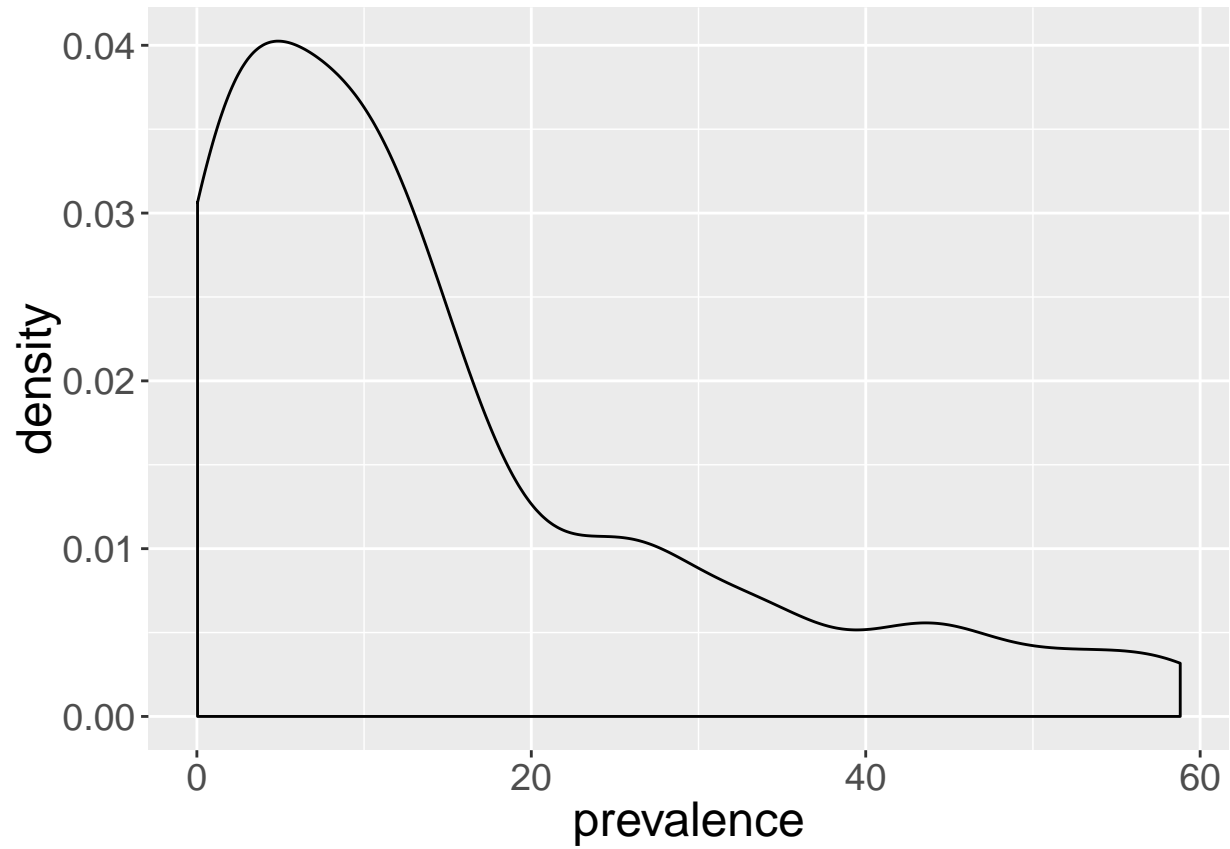
ggplot(alldat,aes(x=prevalence)) + geom_histogram()
```



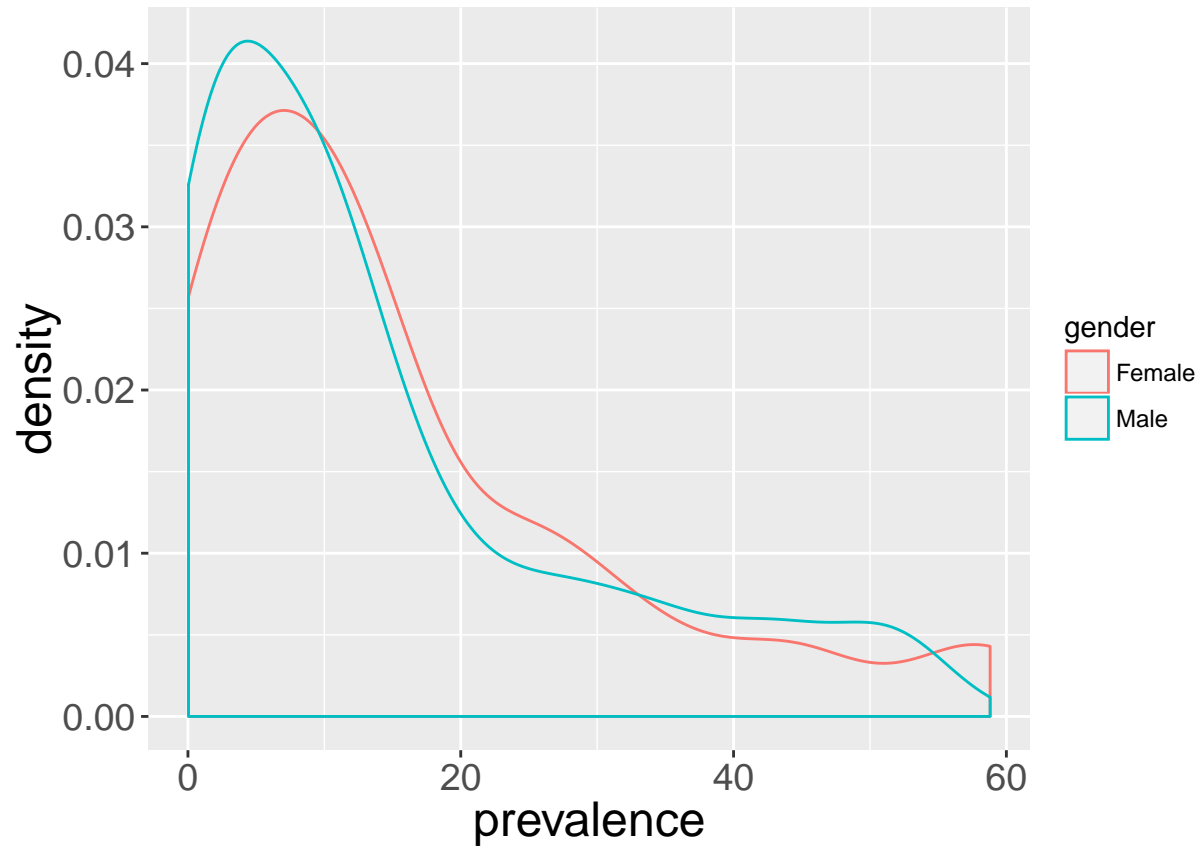
```
ggplot(alldat, aes(x=prevalence)) + geom_histogram(binwidth=5)
```



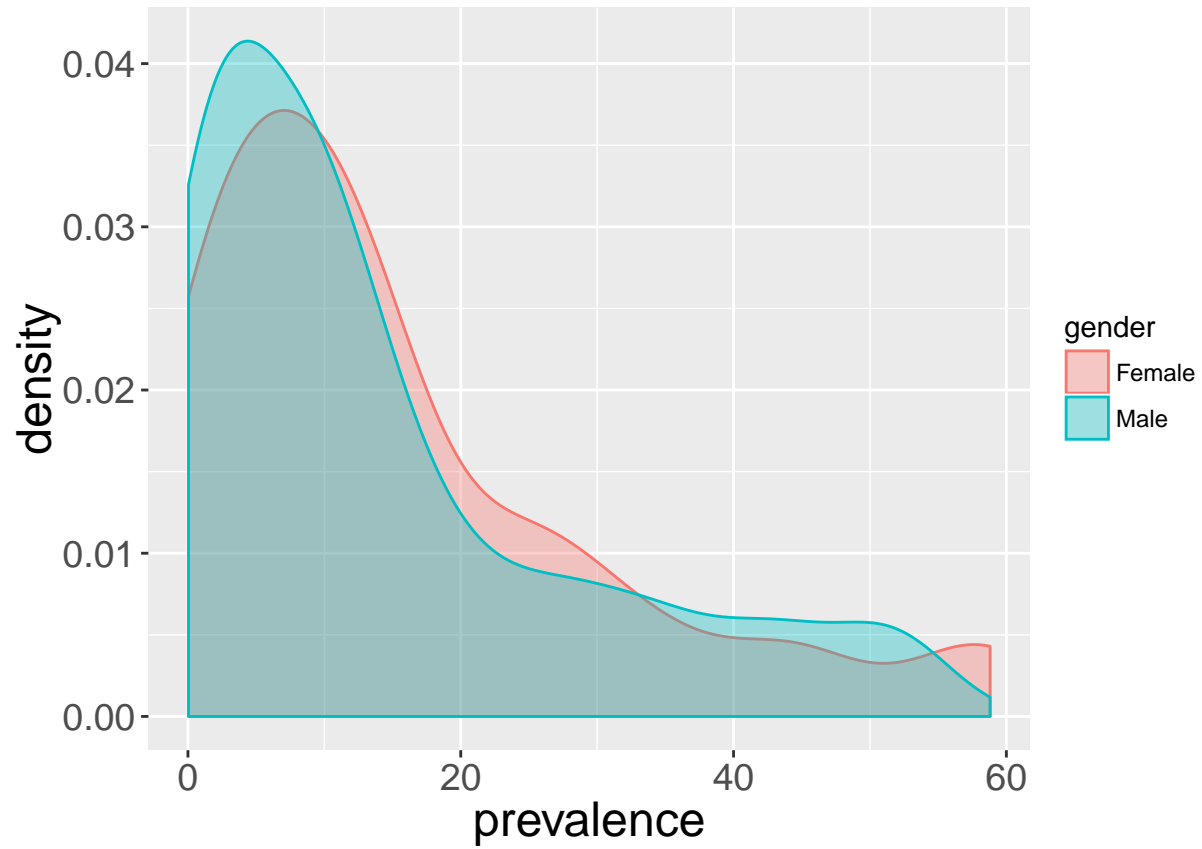
```
ggplot(alldat, aes(x=prevalence)) + geom_density()
```



```
ggplot(alldat, aes(x=prevalence, color=gender)) + geom_density()
```

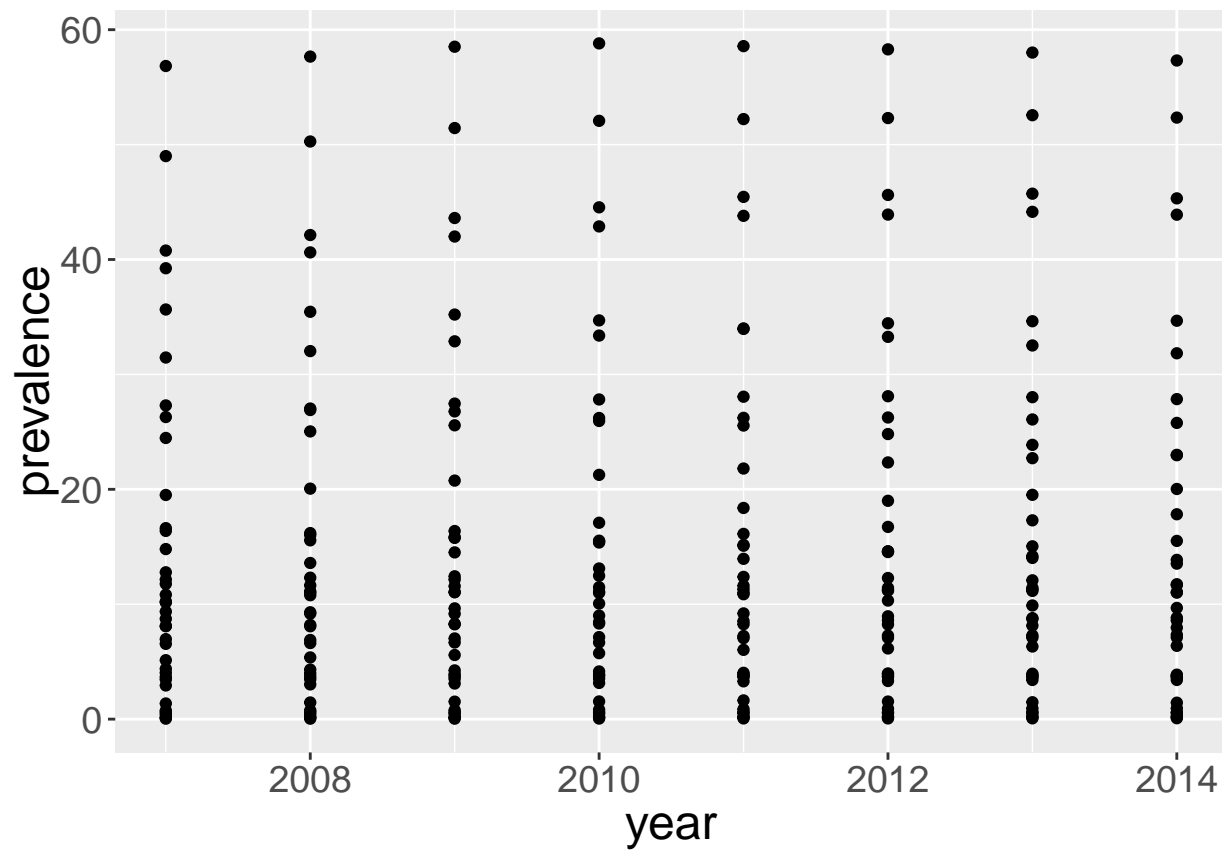


```
ggplot(alldat, aes(x=prevalence, color=gender, fill=gender)) + geom_density(alpha=0.35)
```

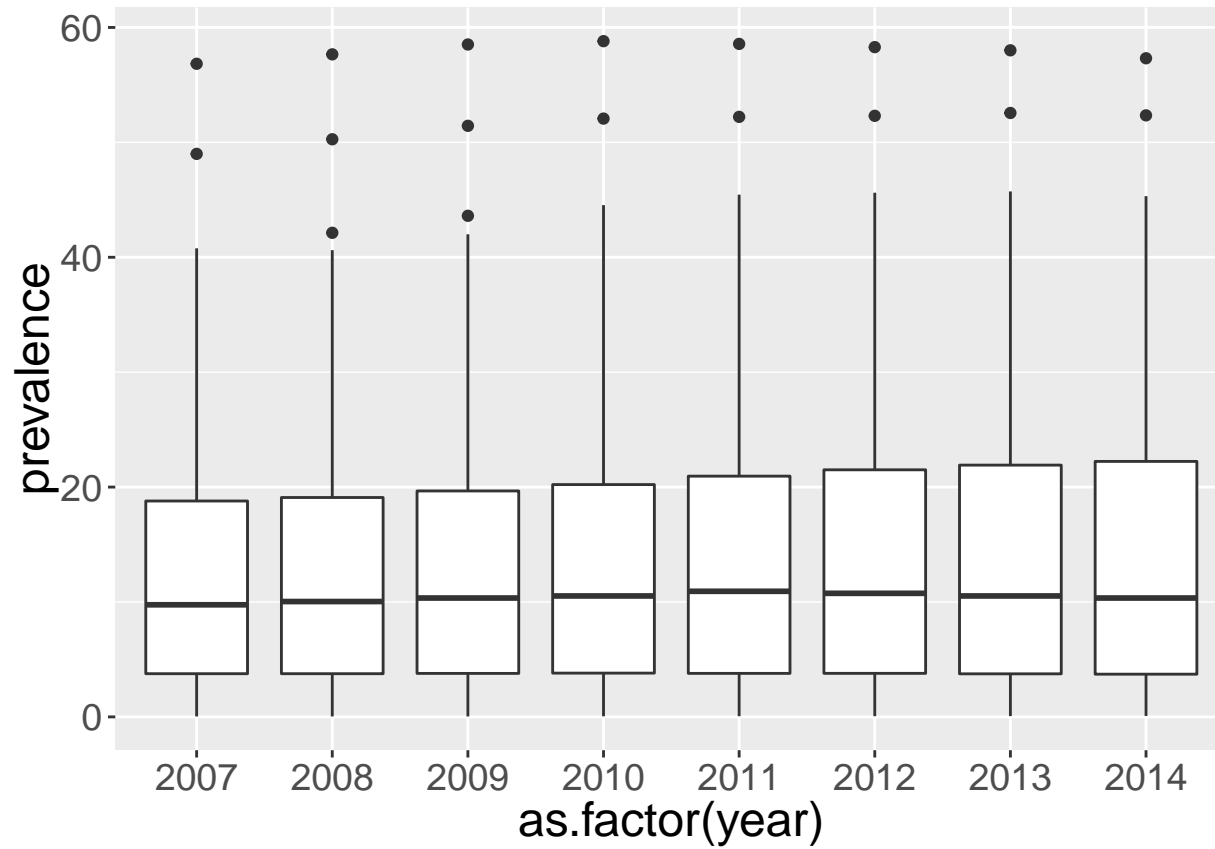


Geoms (Two Variables)

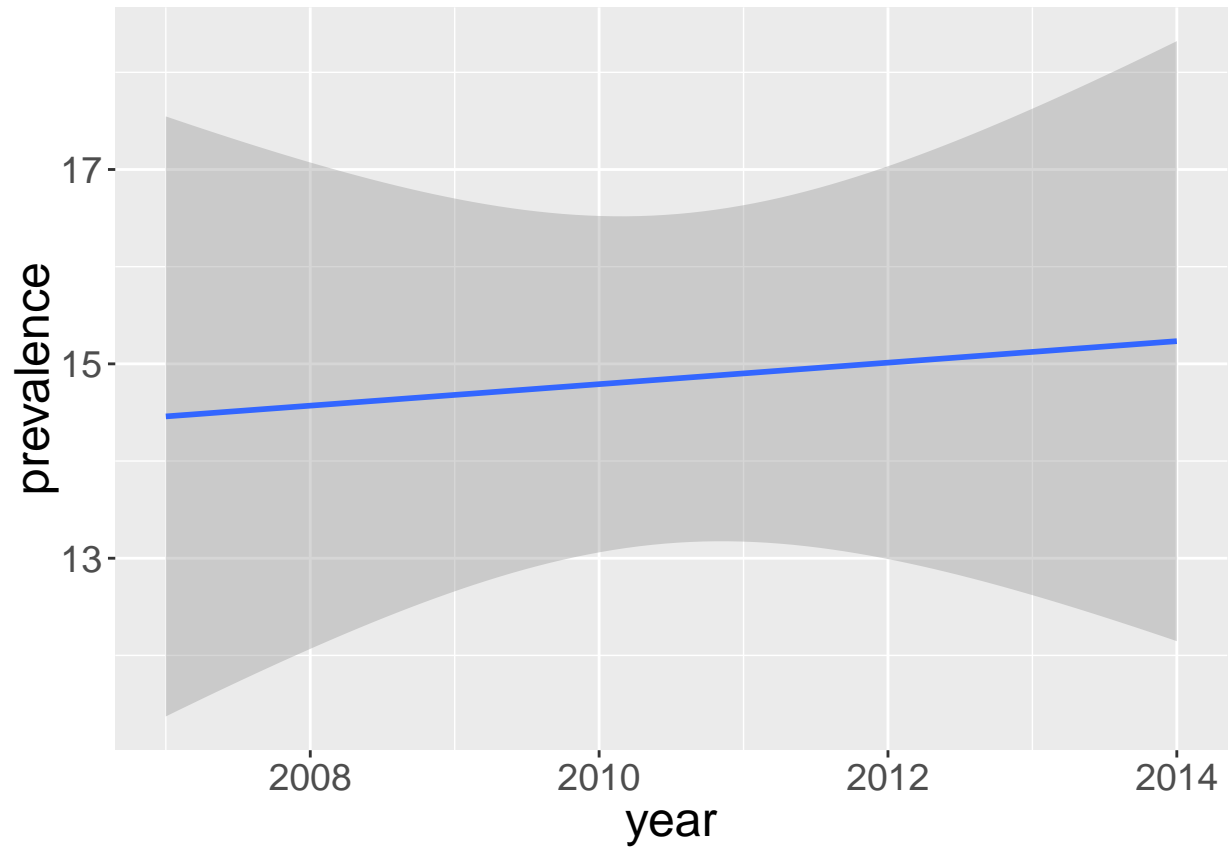
```
ggplot(alldat, aes(x=year, y=prevalence)) + geom_point()
```



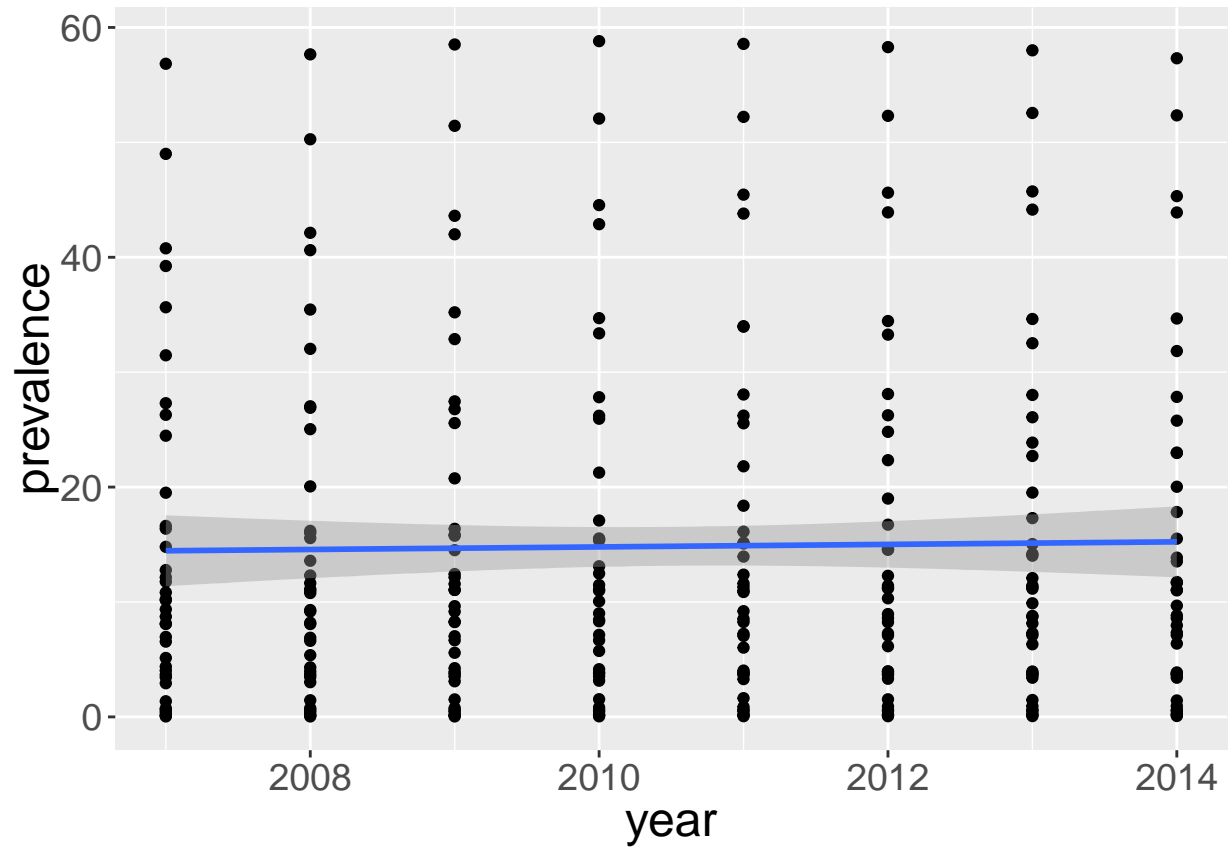
```
ggplot(alldat, aes(x=as.factor(year), y=prevalence)) + geom_boxplot()
```



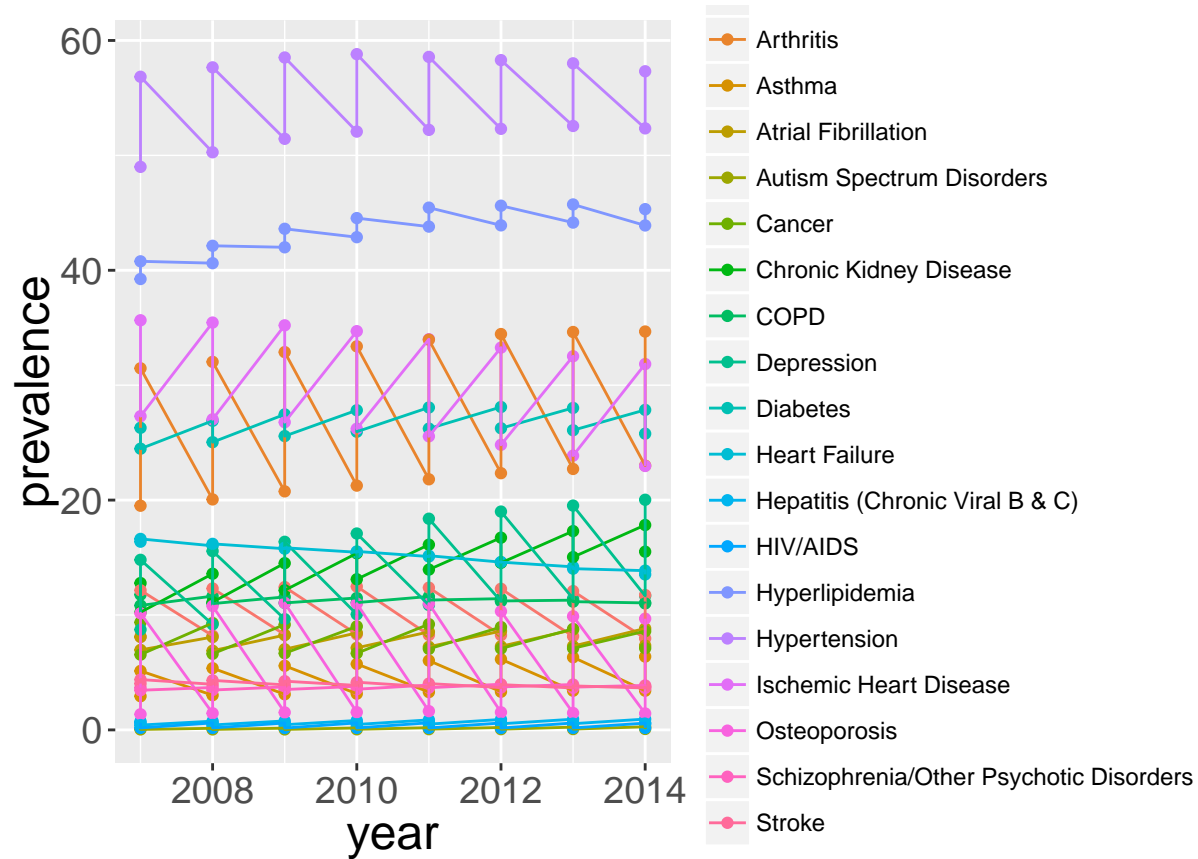
```
ggplot(alldat, aes(x=year, y=prevalence)) + geom_smooth(method=lm)
```



```
ggplot(alldat, aes(x=year, y=prevalence)) + geom_point() + geom_smooth(method=lm)
```

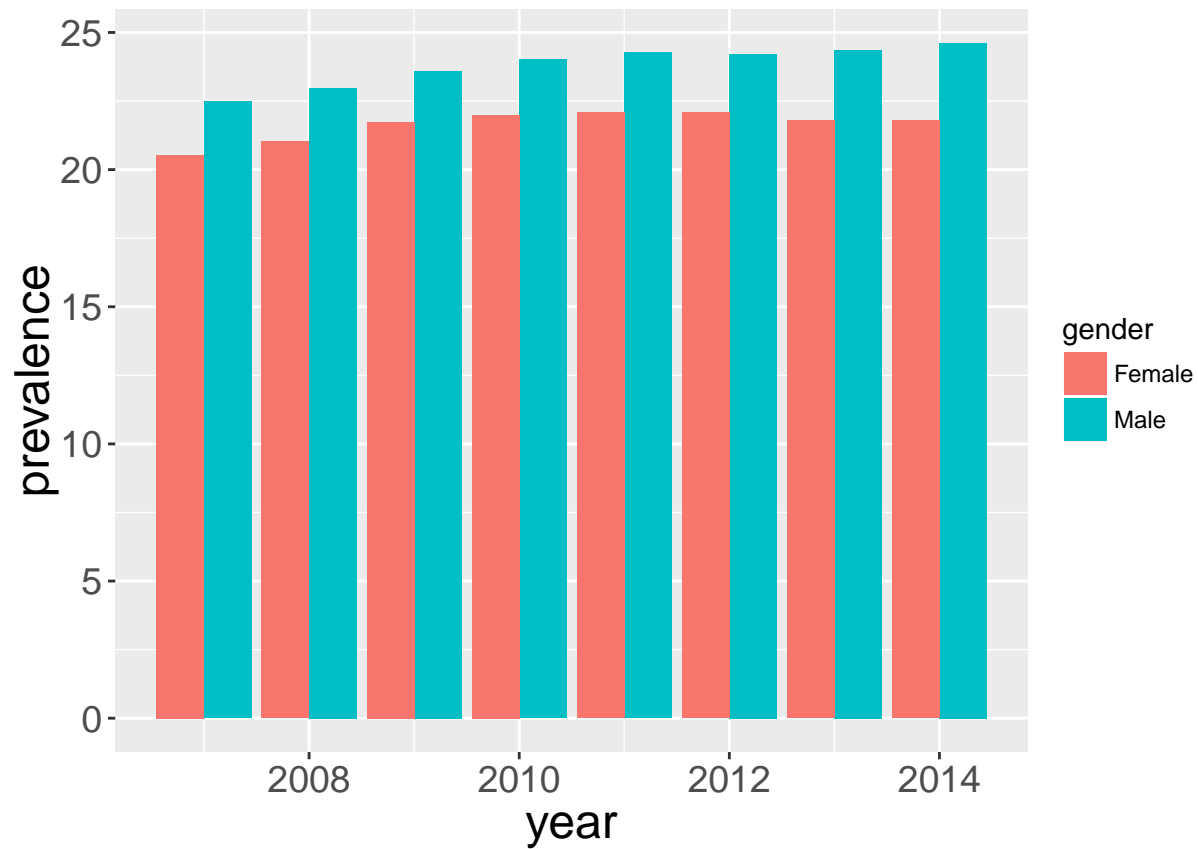


```
ggplot(alldat, aes(x=year, y=prevalence, color=chronicCondition)) + geom_point() + geom_line()
```

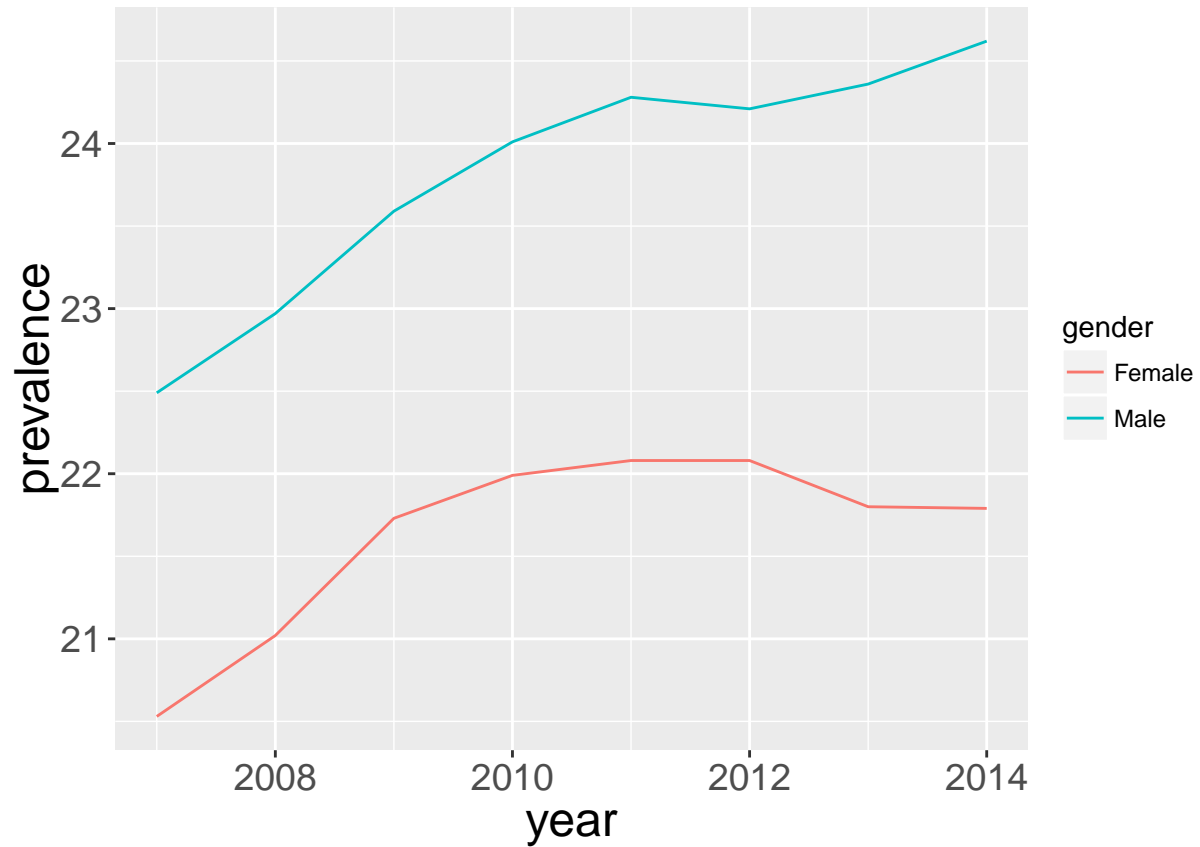


Geoms (Two Variables) cont.

```
ut.diabetes <- ccdat[ccdat$state=="Utah" &  
  ccdat$age=="All" &  
  ccdat$chronicCondition=="Diabetes",]  
  
ggplot(ut.diabetes,aes(x=year,y=prevalence,fill=gender)) + geom_bar(stat="identity",position="dodge")
```

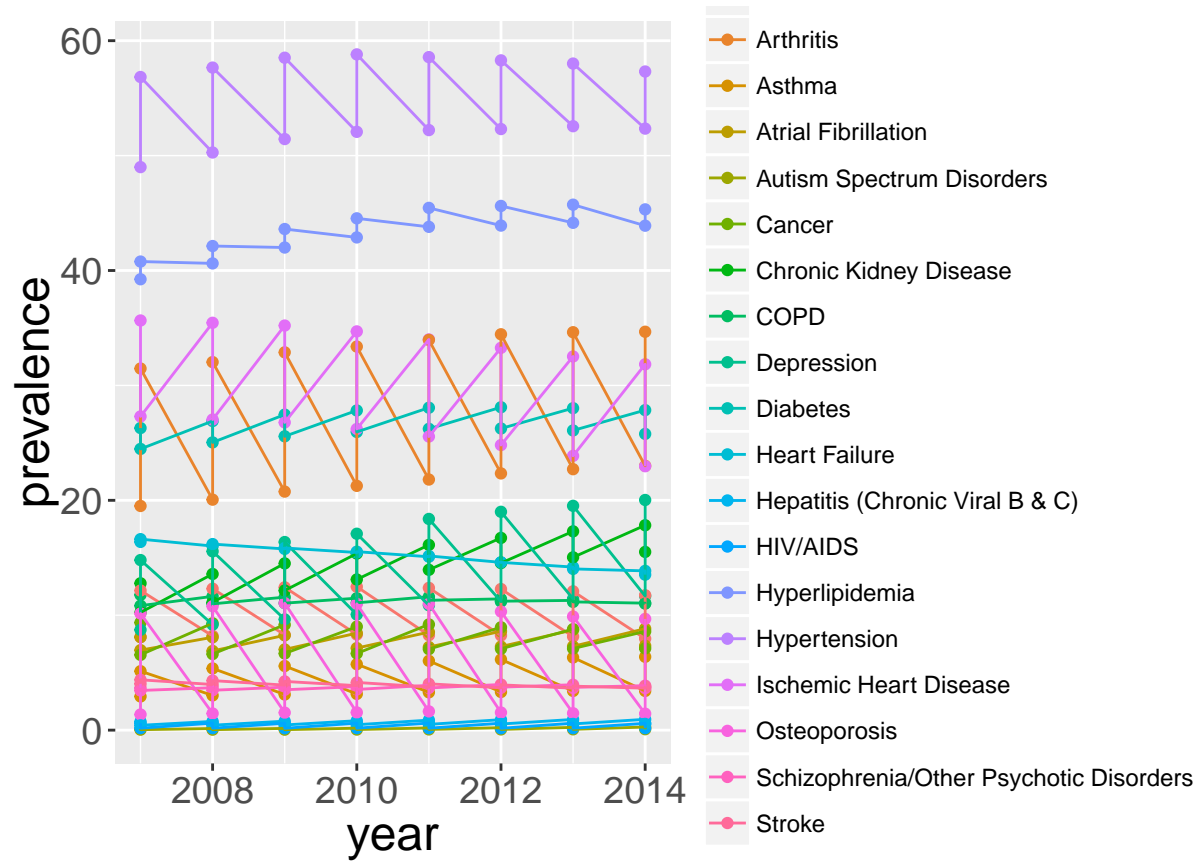


```
ggplot(ut.diabetes,aes(x=year,y=prevalence,color=gender)) + geom_line()
```

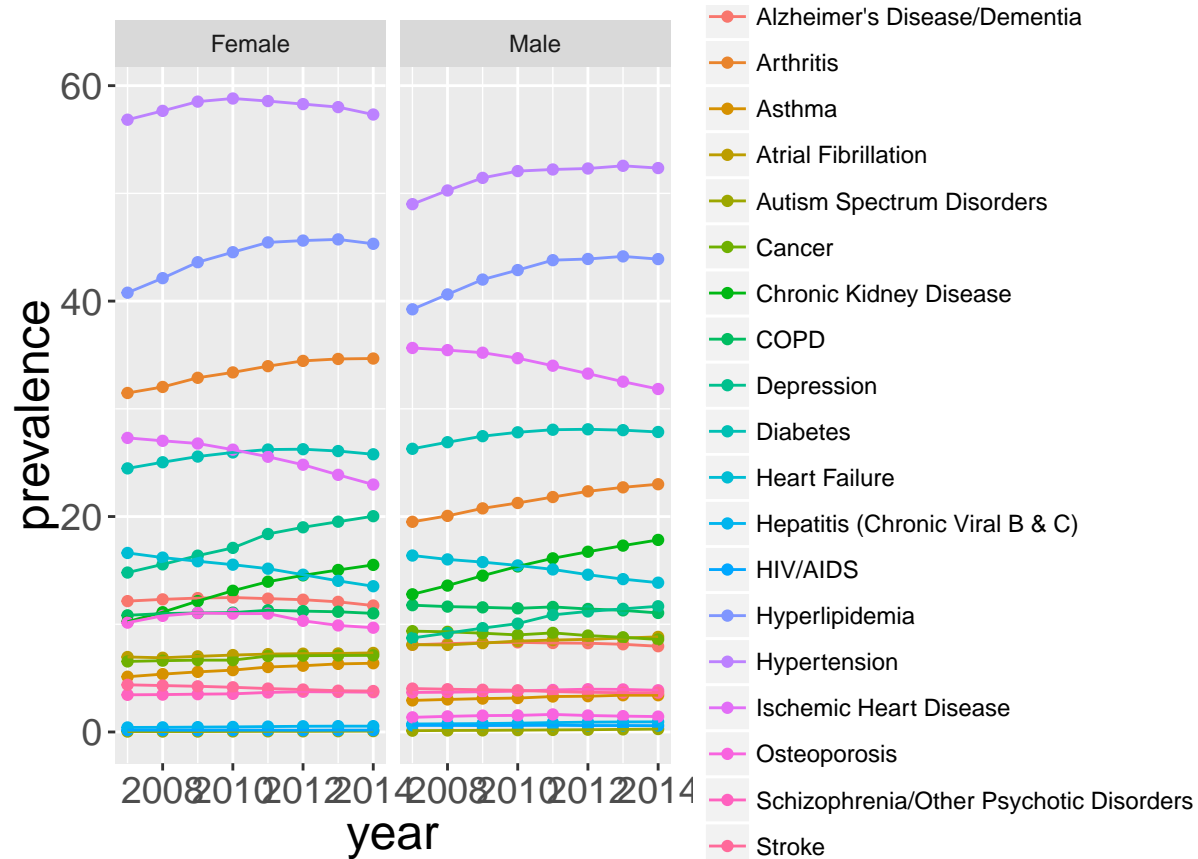


Faceting

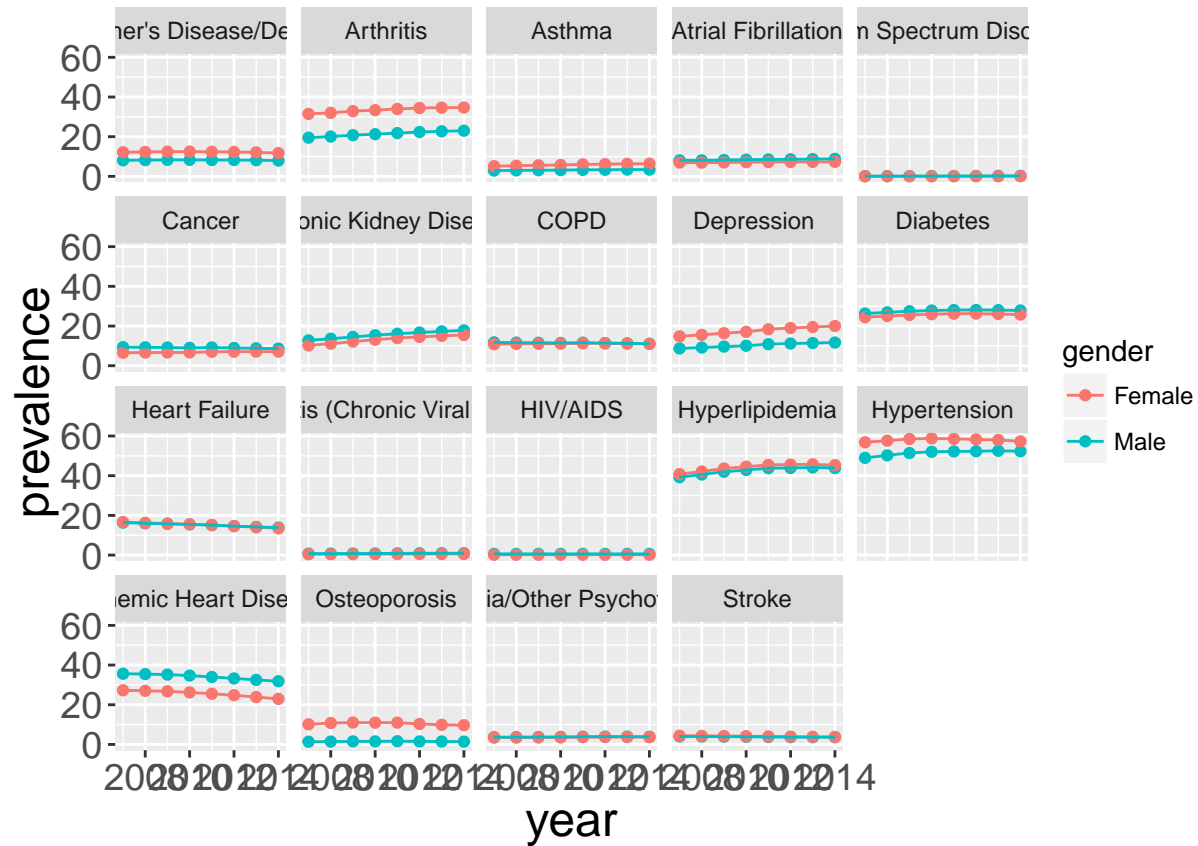
```
ggplot(alldat, aes(x=year, y=prevalence, color=chronicCondition)) + geom_point() + geom_line()
```



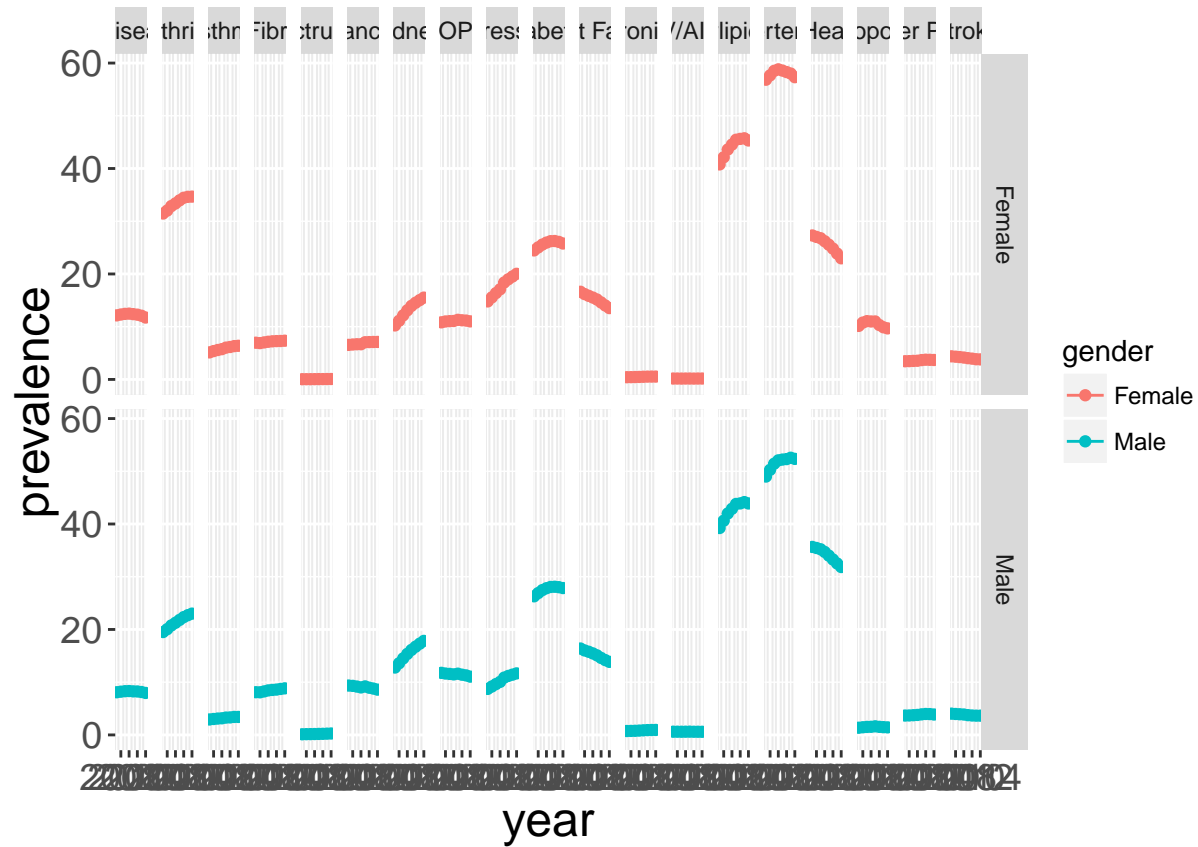

```
ggplot(alldat, aes(x=year, y=prevalence, color=chronicCondition)) + geom_point() + geom_line() + facet_wrap(~gender)
```



```
ggplot(alldat, aes(x=year, y=prevalence, color=gender)) + geom_point() + geom_line() + facet_wrap(~chronicCondition)
```



```
ggplot(alldat, aes(x=year, y=prevalence, color=gender)) + geom_point() + geom_line() + facet_grid(gender~chronicCondition)
```



Citation

- ggplot2 Cheat Sheet
- CMS Medicare Chronic Conditions data from cms.gov
- R Graphics Cookbook by Winston Chang

Questions?