2 - Advanced Graphics in R

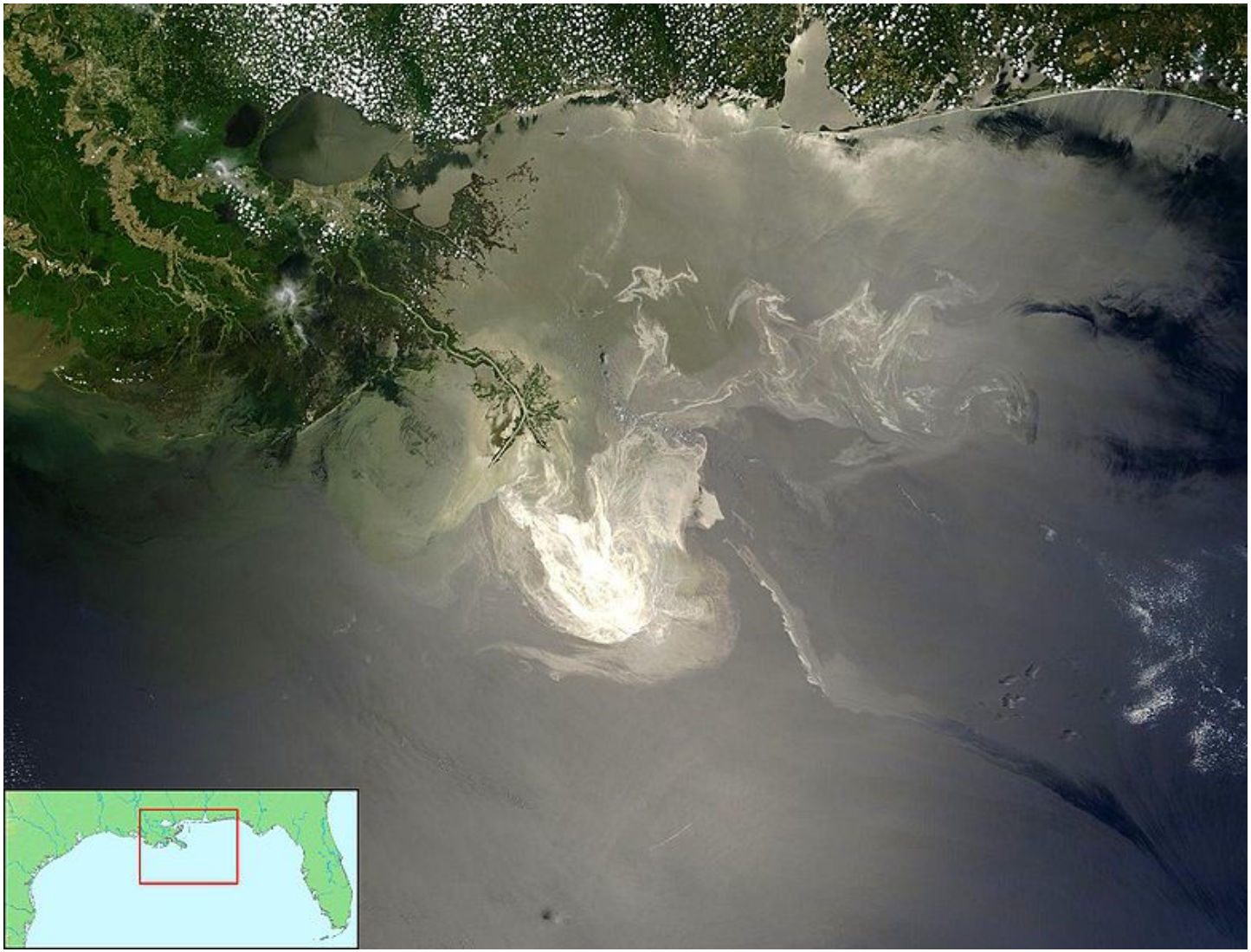
03 - Plotting Using Layers

Iowa State University

Outline

- Data Sources
- Layers
- `ggplot()` vs. `qplot()`

Deepwater Horizon Oil Spill



Data Sets

NOAA Data

- National Oceanic and Atmospheric Administration
- Temperature and Salinity Data in Gulf of Mexico
- Measured using Floats, Gliders and Boats

US Fisheries and Wildlife Data

- Animal Sightings on the Gulf Coast
- Birds, Turtles and Mammals
- Status: Oil Covered or Not

Both data sets have geographic coordinates for ever observation

Loading NOAA Data

NOAA data is a `.rdata` file so we need to read it in specially

- Download the data from http://www.public.iastate.edu/~hofmann/looking-at-data/data/noaa.rdata
- Save the noaa data file from the website to your working directory folder
- To figure out your working directory location use `getwd()`
- Then use the code below to load the data into R

```
  setwd(" - your WD location here - ")
  load("noaa.rdata")
options(width=65)
  ls()
```

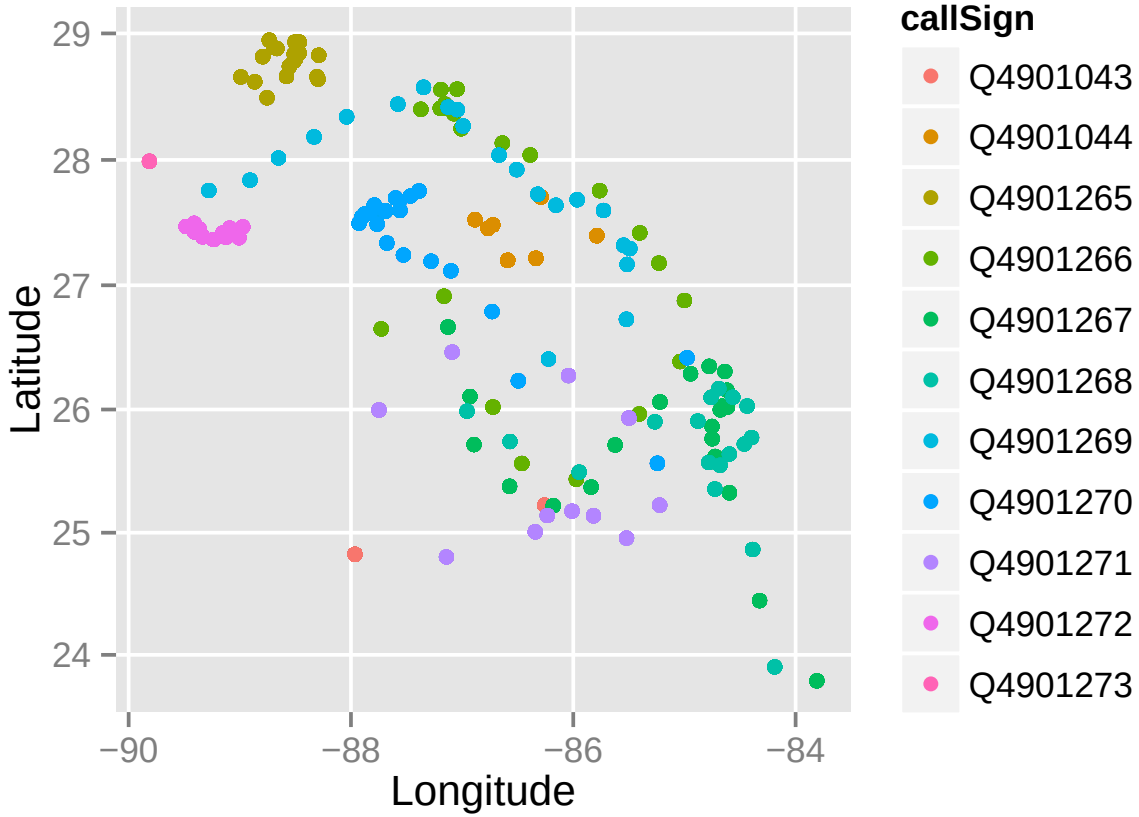
```
## [1] "animals" "boats"   "floats"  "gliders" "rig"     "states"
```

Floats

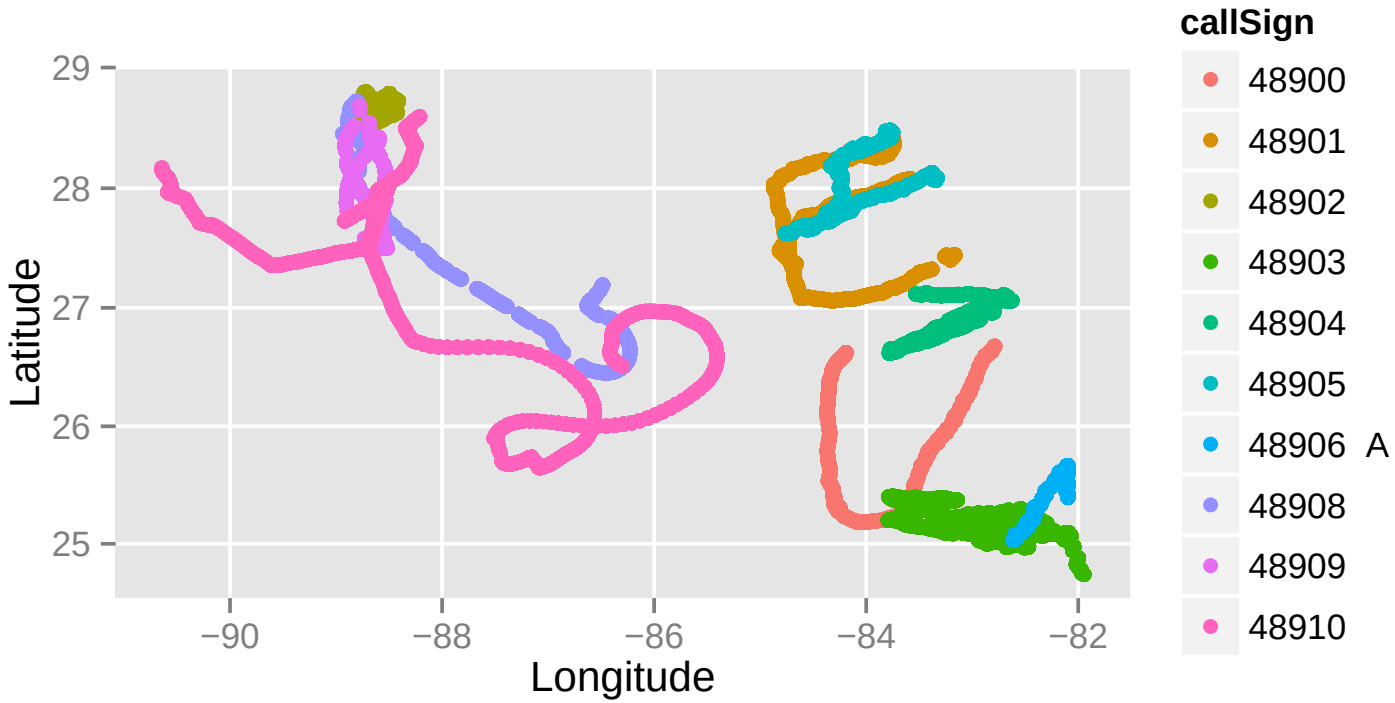
Lets take a peek at the top of the floats NOAA data.

```
##   callSign Date_Time JulianDay Time_QC Latitude
## 1 Q4901043 7/12/2010   2455390       1    24.82
## 2 Q4901043 7/12/2010   2455390       1    24.82
## 3 Q4901043 7/12/2010   2455390       1    24.82
##   Longitude Position_QC Depth Depth_QC Temperature
## 1    -87.96           1     2        1       29.83
## 2    -87.96           1     4        1       29.65
## 3    -87.96           1     6        1       29.53
##   Temperature_QC Salinity Salinity_QC  Type
## 1              1    36.59           1 Float
## 2              1    36.58           1 Float
## 3              1    36.58           1 Float
```

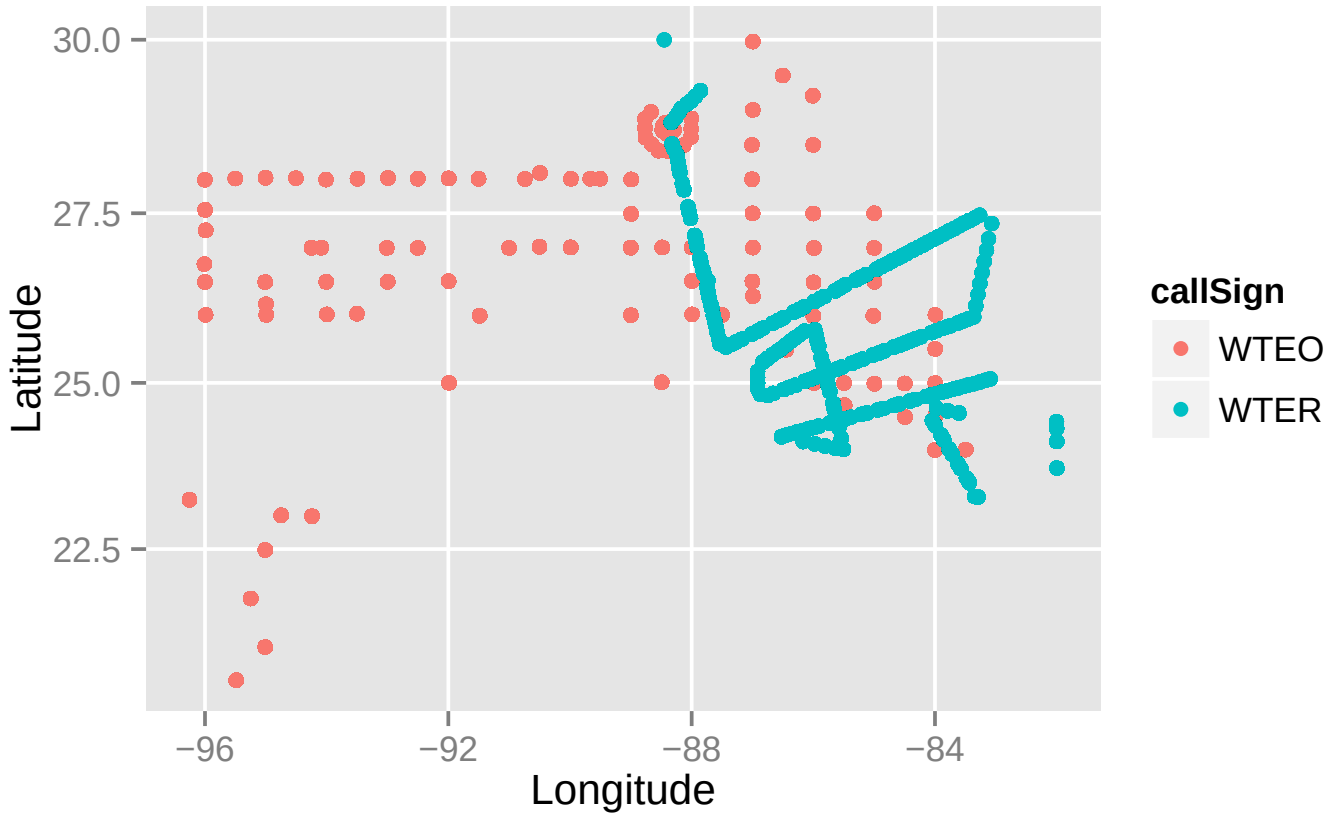
Floats


Latitude
29
28
27
26
25
24
−90
−88
−86
−84
Longitude
callSign
Q4901043
Q4901044
Q4901265
Q4901266
Q4901267
Q4901268
Q4901269
Q4901270
Q4901271
Q4901272
Q4901273

Gliders


Latitude
29
28
27
26
25
−90
−88
−86
−84
−82
Longitude
callSign
48900
48901
48902
48903
48904
48905
48906 A
48908
48909
48910

Boats


30.0
27.5
25.0
22.5
Latitude
−96
−92
−88
−84
Longitude
callSign
WTEO
WTER

Layering

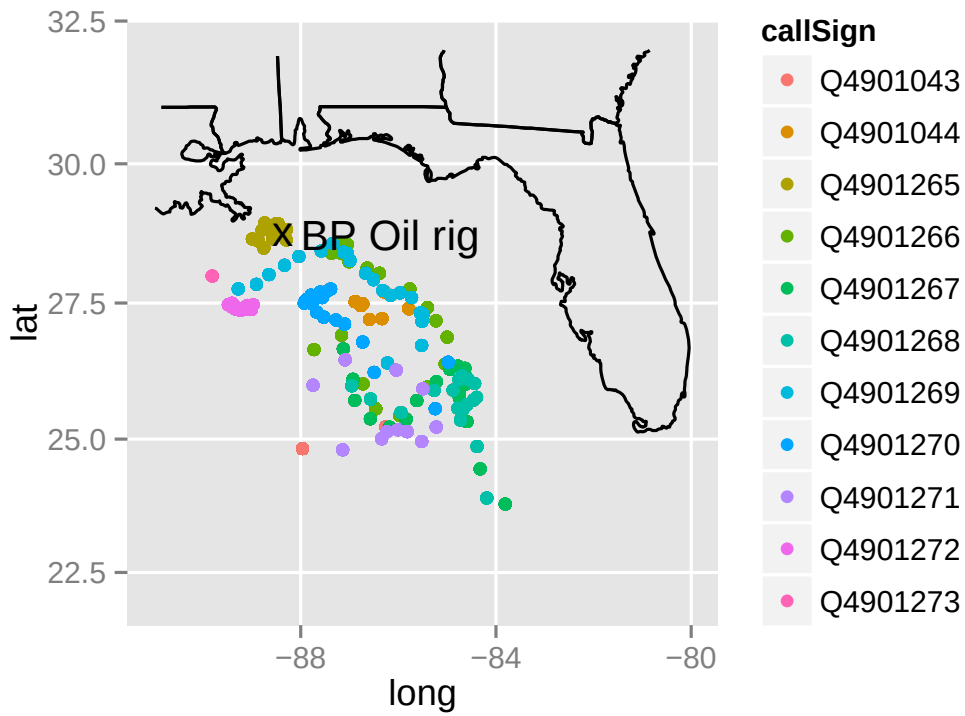
This data has the same context - a common time and common place

- Want to aggregate information from different sources onto a common plot
- Start with a common background the lat/long grid
- With `ggplot2` we will superimpose data onto this grid in layers

Layers

... to give you an idea ...

```
ggplot() +    # plot without a default data set
geom_path(data=states, aes(x=long, y=lat, group=group)) +
geom_point(data=floats, aes(x=Longitude, y=Latitude, colour=callSign)) +
geom_point(aes(x, y), shape="x", size=5, data=rig) +
geom_text(aes(x, y), label="BP Oil rig", shape="x",
          size=5, data=rig, hjust = -0.1) +
xlim(c(-91, -80)) + ylim(c(22,32)) + coord_map()
```



Layering

- Most maps (and many plots) have multiple layers of data. The layers may be from the same or different datasets.
- ggplot2 build around this same idea. Very easy to add additional layers to the plot. To do this we need to understand a little more about the underlying theory.

What is a Plot?

Any plot is composed of:

1. A default dataset
2. A coordinate system
3. layers of geometric objects (geoms)
4. A set of aesthetic mappings (taking information from the data and converting into an attribute of the plot)
5. A scale for each aesthetic
6. A facetting specification (multiple plots based on subsetting the data)

Floats Decomposed

Data: floats
Mappings:

x	=	Longitude
y	=	Latitude
color	=	CallSign

Layers:

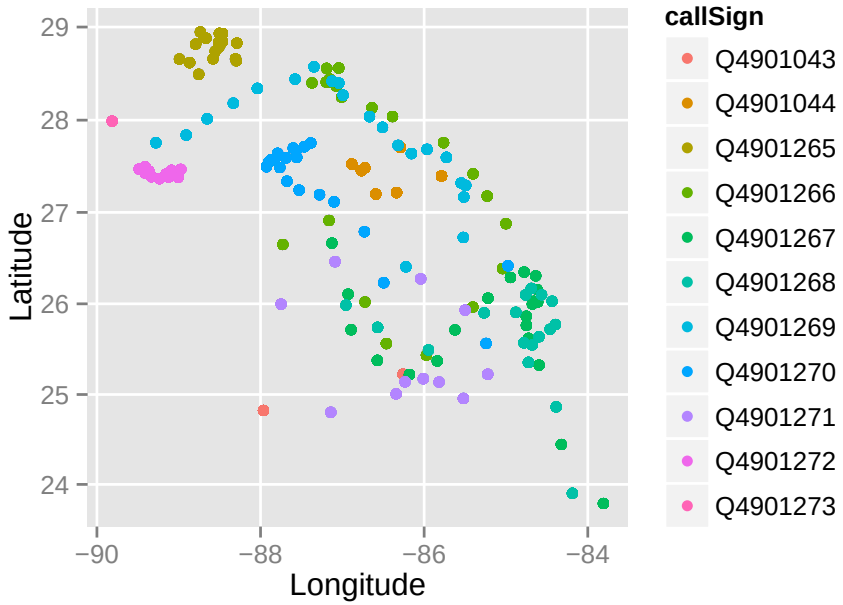
Geoms: Points

Scales:

x & y position

discrete color

Faceting: None



`qplot()` vs. `ggplot()`

`qplot()` stands for "quickplot"

- automatically chooses default settings to make life easier
- less control over plot construction

`ggplot()` stands for "grammar of graphics plot"

- Constructs the plot using components listed in previous slides

qplot() vs. ggplot()

Two ways to construct the same plot for float locations

```
qplot(Longitude, Latitude, colour=callSign, data=floats)

###

ggplot(data=floats,
       aes(x=Longitude, y=Latitude, colour=callSign)) +
  geom_point() +
  scale_x_continuous() +
  scale_y_continuous() +
  scale_colour_discrete ()

###

# But we don't need to be quite so verbose.  Scales are
# added automatically and first two aes params are x and y:
ggplot(floats,
       aes(Longitude, Latitude, colour = callSign)) +
  geom_point()
```

Floats Decomposed

Data: floats
Mappings:

x	=	Longitude
y	=	Latitude
color	=	CallSign

Layers:

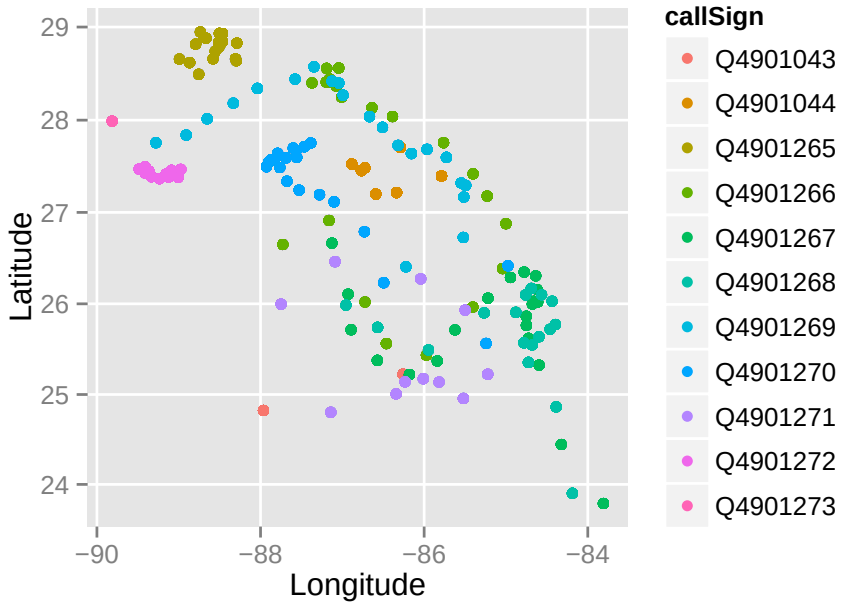
Geoms: Points

Scales:

x & y position

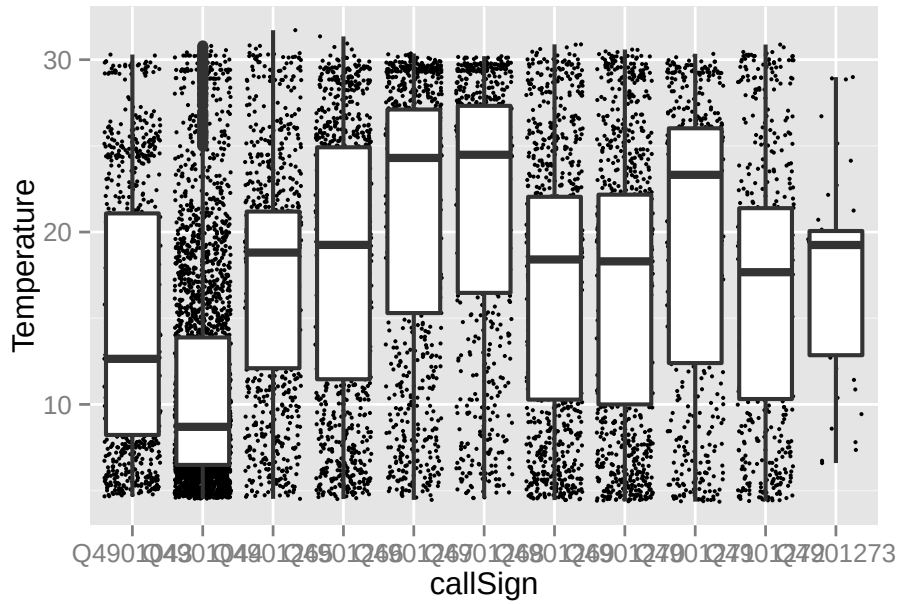
discrete color

Faceting: None



qplot() vs. ggplot()

Data: floats
Mappings:
x = CallSign
y = Temperature
Layers:
Geoms: Jittered Points
Boxplots
Scales:
x & y position
Faceting: None



qplot() vs. ggplot()

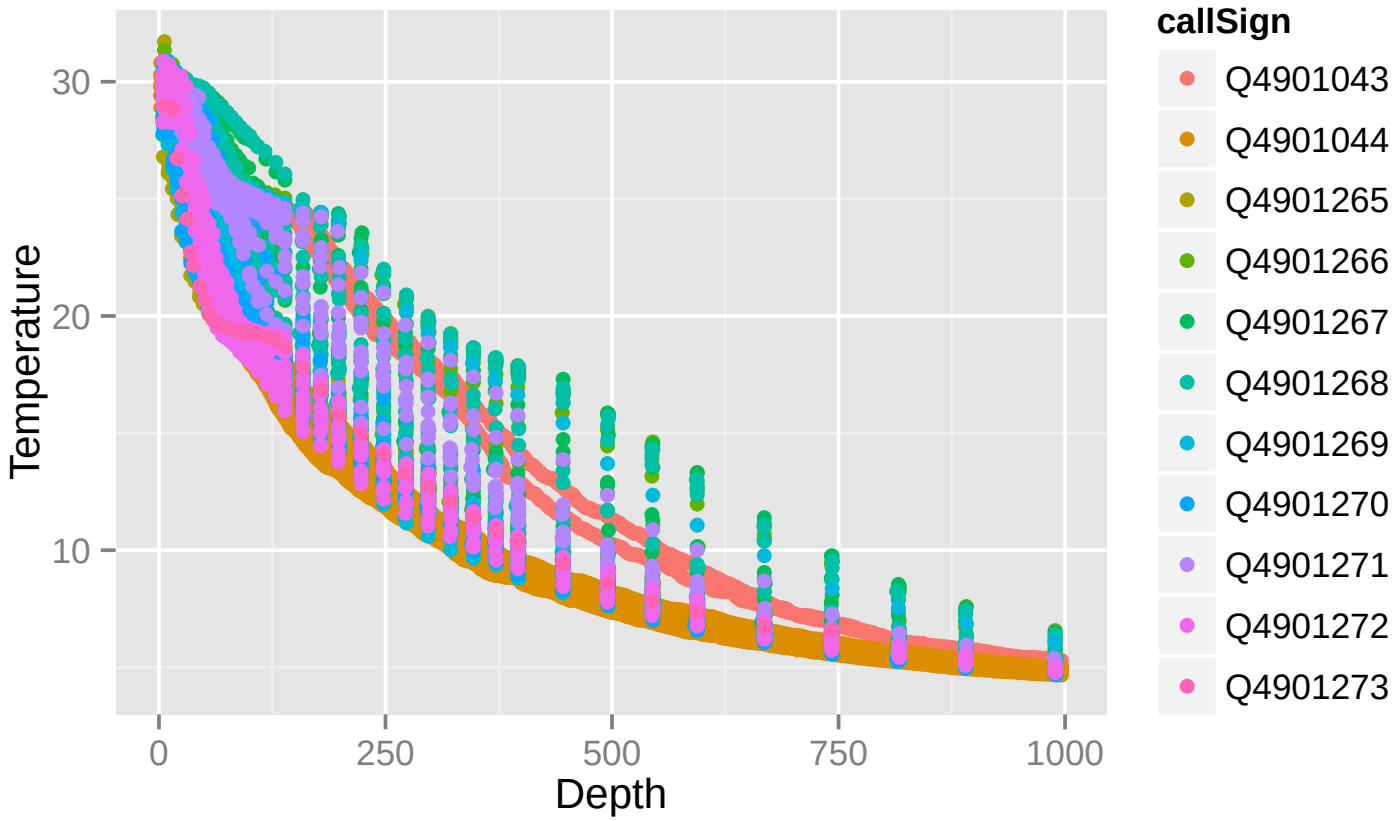
Again, there are two ways to construct this plot

```
# Temperature by callSign
### using qplot
qplot(callSign, Temperature, data=floats,
      geom=c("jitter", "boxplot"))

### using ggplot
ggplot(floats, aes(callSign, Temperature)) +
  geom_jitter() +
  geom_boxplot()
```

Your Turn

Find the `ggplot()` statement that creates this plot



What is a layer?

A layer added to `ggplot()` can be a geom ...

- the type of geometric object
- the statistic mapped to that object
- the data set from which to obtain the statistic

... or a position adjustment to the scales

- Changing the axes scale
- Changing the color gradient

Layer Examples

Plot	Geom	Stat
Scatterplot	point	identity
Histogram	bar	bin count
Smoother	line + ribbon	smoother function
Binned Scatterplot	rectangle + color	2d bin count

More geoms described at http://docs.ggplot2.org/current/

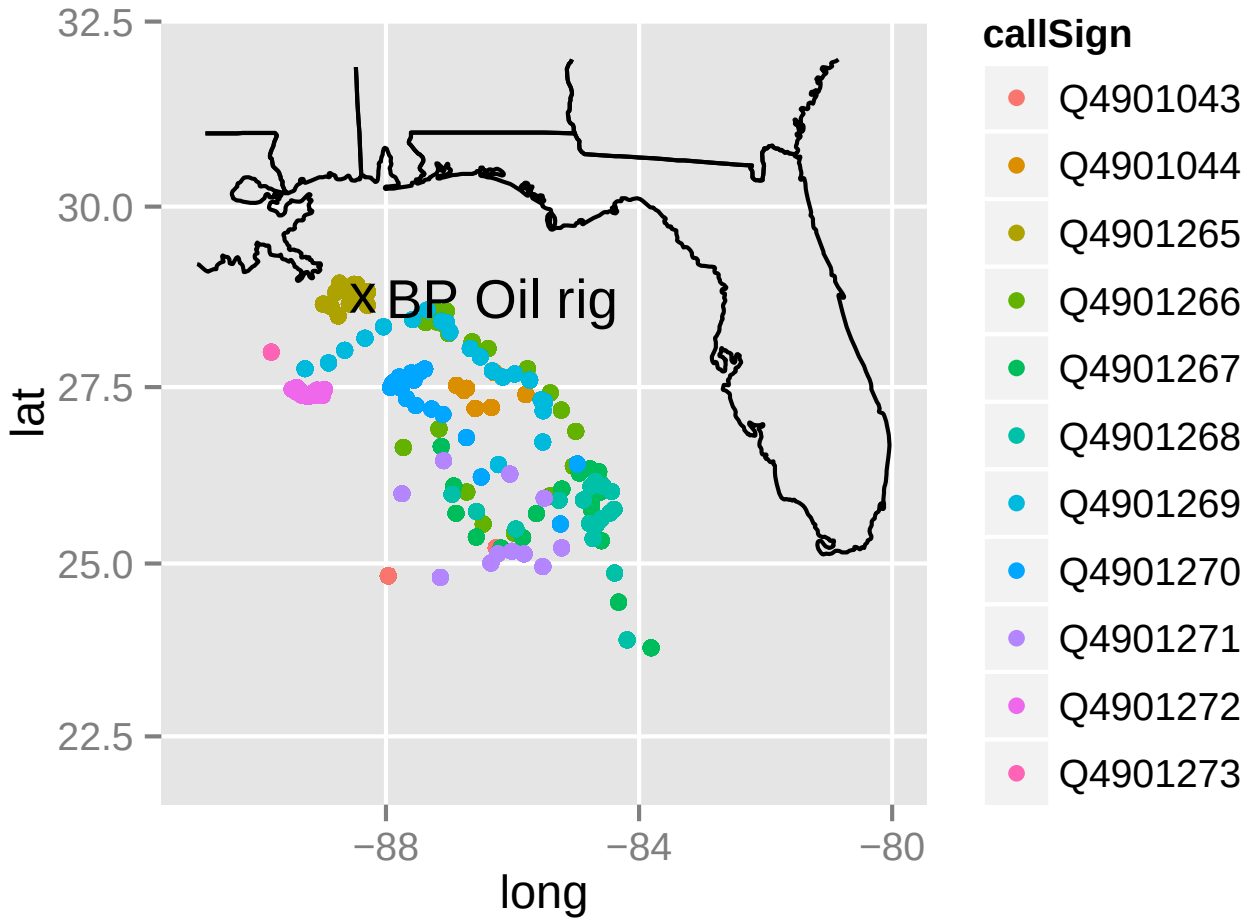
Piecing things together

Want to build a map using NOAA data

- Coordinate system (mapping Long-Lat to X-Y)
- Add layer of state outlines
- Add layer of points for float locations
- Add layers for Oil Rig marker and label
- Adjust the range of x and y scales

```
ggplot() +    # plot without a default data set
 geom_path(data=states, aes(x=long, y=lat, group=group)) +
 geom_point(data=floats,
            aes(x=Longitude, y=Latitude, colour=callSign)) +
 geom_point(aes(x, y), shape="x", size=5, data=rig) +
 geom_text(aes(x, y), label="BP Oil rig", shape="x",
       size=5, data=rig, hjust = -0.1) +
 xlim(c(-91, -80)) +
 ylim(c(22,32))
```

Piecing things together



Your Turn

- Read in the animal.csv data
- Plot the location of animal sightings on a map of the region
- On this plot try to color points by class of animal and/or status of animal
- Advanced: Could we indicate time somehow?