pollution-report

Omer Cofman

April 27, 2016

This is a global dataset of 1571 locations where surface manta tows were conducted.

Samples were divided into 4 size categories. Weights and particle counts were recorded for each category.

Since the 4th categories of both weight and particle counts were mostly NA values, they were gone in the data cleaning operation.

More attributes I cleaned out are Date, Sea State, Source (scientist), Info and Comments.

We are left with the Latitude, Longitude, and 3 size categories by weights and particle counts.

Attribute Information:

1. Latitude
2. Longitude
3. CD1: numeric, particle counts for debris between .335-.999 mm, units: [1/kmˆ2]
4. CD2: numeric, particle counts for debris between 1.00-4.75 mm, units: [1/kmˆ2]
5. CD3: numeric, particle counts for debris between 4.75-200 mm, units: [1/kmˆ2]
6. WD1: numeric, particle weight for debris between .335-.999 mm, units: [g/kmˆ2]
7. WD2: numeric, particle weight for debris between 1.00-4.75 mm, units: [g/kmˆ2]
8. WD3: numeric, particle weight for debris between 4.75-200 mm, units: [g/kmˆ2]

Let's see some data

```
head(data);
```

```
##   X Latitude Longitude      cd1      cd2      cd3  wd1   wd2     wd3
## 1 2  19.9432  -64.5649 58102.96 22326.06  2226.17 4.45 26.83    4.23
## 2 3  20.2173  -64.3828  6639.79   573.26  1067.11 1.04 28.69   40.79
## 3 4  20.4521  -64.1968 15246.71 11841.41   991.66 2.57 42.62 3503.27
## 4 5  21.1293  -63.8333  5347.35     0.00  1420.39 1.15 12.86    4.26
## 5 6  21.4730  -63.5899  4090.58   766.52   409.06 0.48 29.63    4.36
## 6 7  21.7367  -63.4227 44914.59  6731.74 33133.71 6.26 47.98 2051.84
```

As you can see in the next chunk, most of the values are zero

```
summary(data)
```

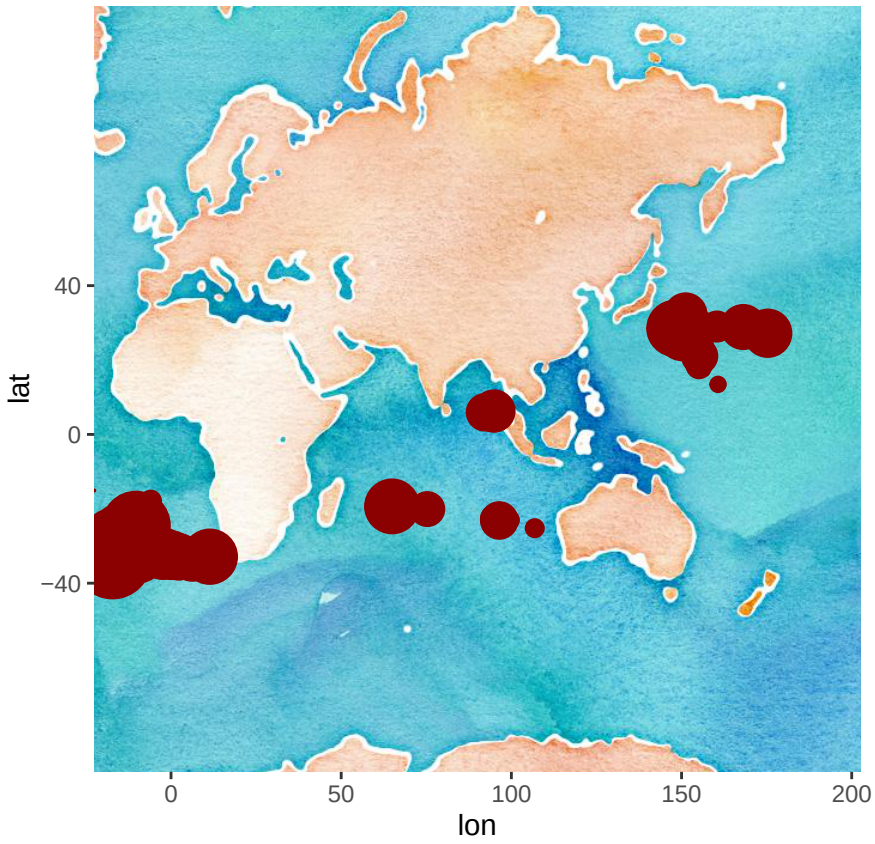
```
##        X             Latitude        Longitude             cd1
##  Min.   :   2.0   Min.   :-63.88   Min.   :-179.994   Min.   :     0.0
##  1st Qu.: 394.5   1st Qu.:-21.76   1st Qu.:-127.242   1st Qu.:     0.0
##  Median : 787.0   Median : 30.26   Median :   6.284   Median :     0.0
##  Mean   : 787.0   Mean   : 13.41   Mean   : -10.210   Mean   :  9612.3
##  3rd Qu.:1179.5   3rd Qu.: 42.46   3rd Qu.:  88.790   3rd Qu.:   882.8
##  Max.   :1572.0   Max.   : 46.65   Max.   : 179.450   Max.   :479454.6
##       cd2               cd3               wd1               wd2
##  Min.   :     0.0   Min.   :     0.0   Min.   :  0.000   Min.   :   0.00
##  1st Qu.:     0.0   1st Qu.:     0.0   1st Qu.:  0.000   1st Qu.:   0.00
```

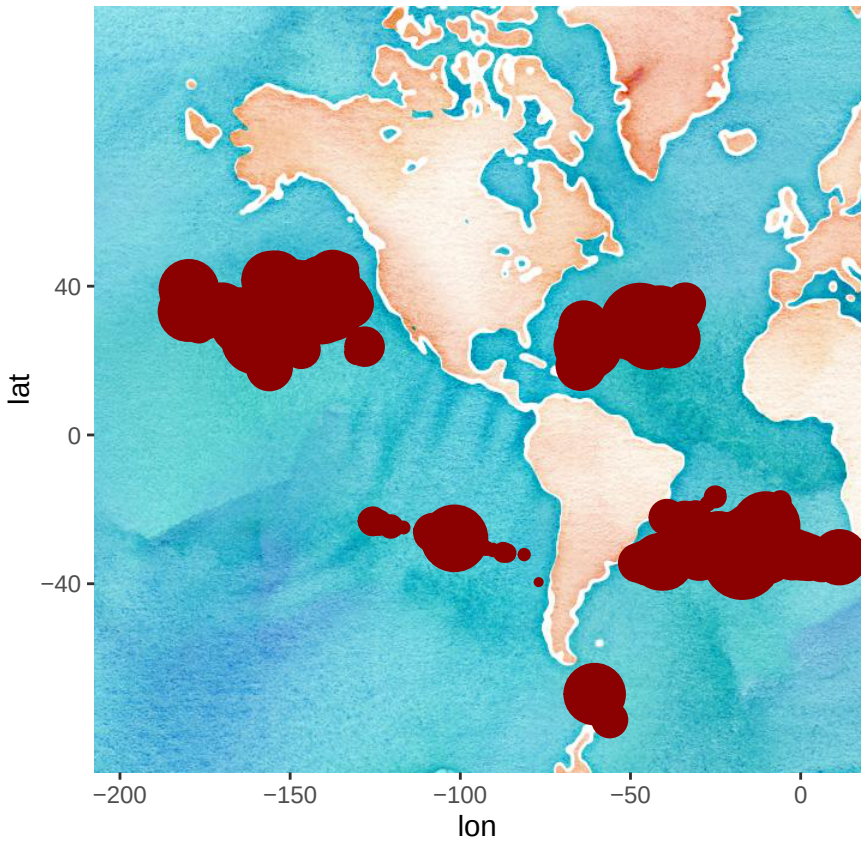
```
##  Median :     0.0   Median :     0.0   Median :  0.000   Median :   0.00
##  Mean   :  9678.3   Mean   :  2158.6   Mean   :  1.853   Mean   :  56.57
##  3rd Qu.:   899.3   3rd Qu.:   831.7   3rd Qu.:  0.000   3rd Qu.:   3.01
##  Max.   :479454.6   Max.   :139777.8   Max.   :296.540   Max.   :9278.23
##       wd3
##  Min.   :    0.00
##  1st Qu.:    0.00
##  Median :    0.00
##  Mean   :  218.62
##  3rd Qu.:   10.36
##  Max.   :25752.82
```

Now we will display the debris on the world map.

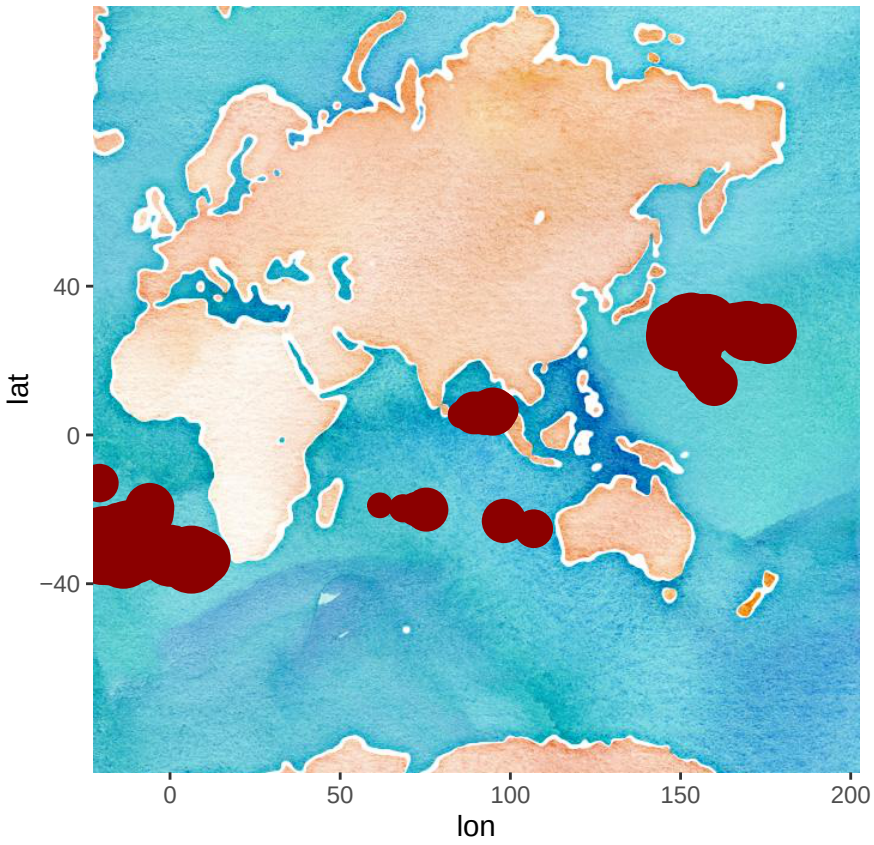
Each couple of maps corresponds to one of the particle counts columns.

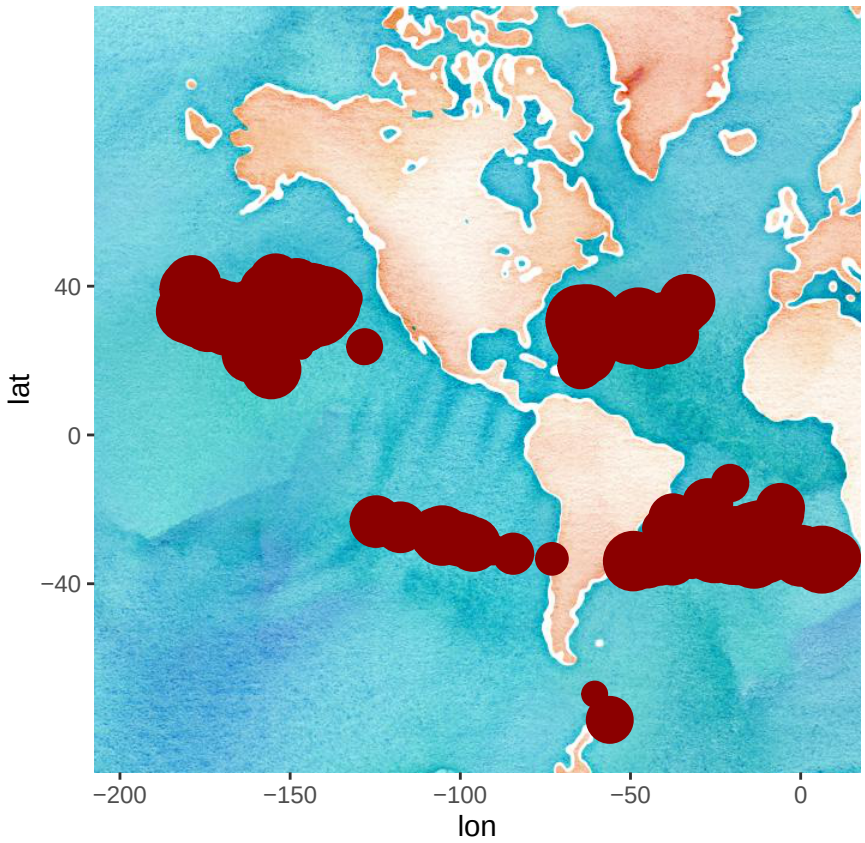
The size of the mark on the map is in proportion to the average size in mm and the weight of the debris CD1*WD1/.677



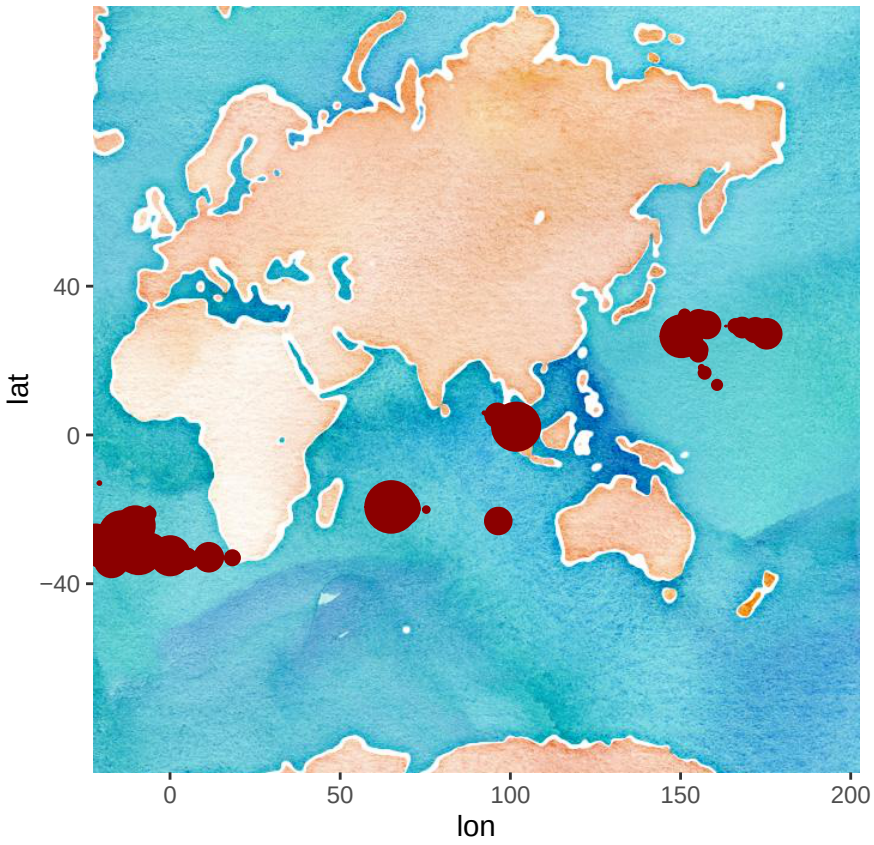


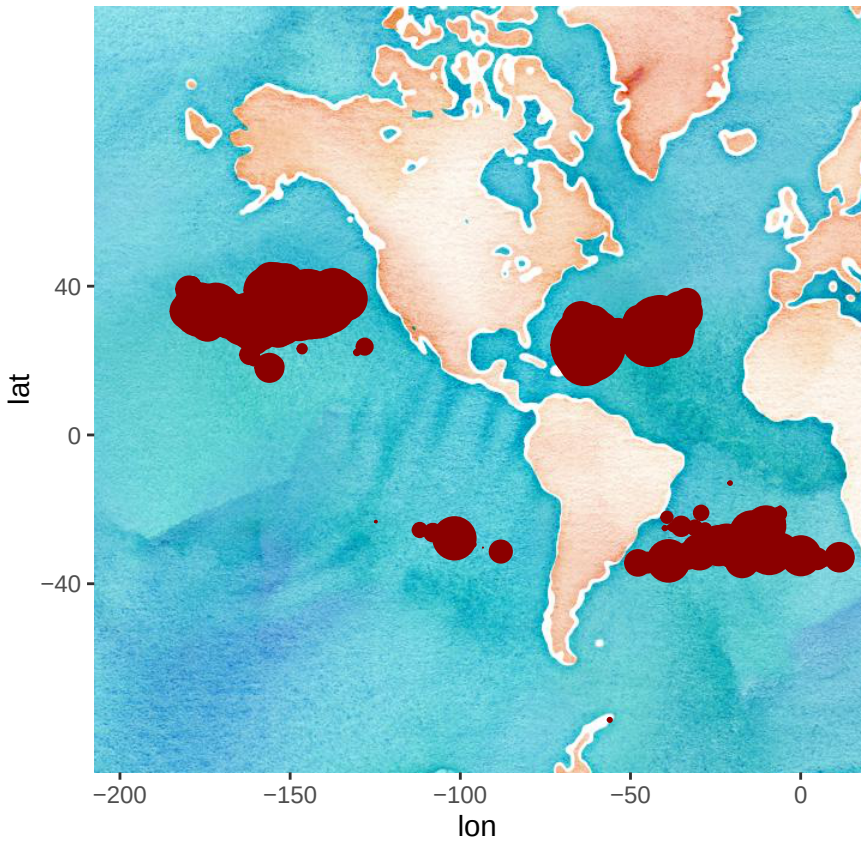
CD2*WD2/2.875


40
0
−40
lat
0
50
100
150
200
lon



CD3*WD3/100


lat
40
0
−40
0
50
100
150
200
lon



We learn that CD2 objects (1.00-4.75mm) are most of the volume of junk in the oceans.

Now lets see what we can learn from splitting the data by longitude

1. from 150 to 200
2. from 50 to 150
3. from 30 to 50
4. from -80 to 30
5. from -200 to -80

```
data.1 = data[data$Longitude<200 & data$Longitude>150,]
data.2 = data[data$Longitude<150 & data$Longitude>50,]
data.3 = data[data$Longitude<50 & data$Longitude>30,]
data.4 = data[data$Longitude<30 & data$Longitude>-80,]
data.5 = data[data$Longitude<-80 & data$Longitude>-200,]
mean(data.1$cd1*data.1$wd1/.677 + data.1$cd2*data.1$wd2/2.875 + data.1$cd3*data.1$wd3/100)
```

```
## [1] 155211.9
```

```
mean(data.2$cd1*data.2$wd1/.677 + data.2$cd2*data.2$wd2/2.875 + data.2$cd3*data.2$wd3/100)
```

```
## [1] 21108.13
```

```
mean(data.3$cd1*data.3$wd1/.677 + data.3$cd2*data.3$wd2/2.875 + data.3$cd3*data.3$wd3/100)
```

```
## [1] NaN
```

```
mean(data.4$cd1*data.4$wd1/.677 + data.4$cd2*data.4$wd2/2.875 + data.4$cd3*data.4$wd3/100)
```

```
## [1] 713673.2
```

```
mean(data.5$cd1*data.5$wd1/.677 + data.5$cd2*data.5$wd2/2.875 + data.5$cd3*data.5$wd3/100)
```

```
## [1] 818267.9
```

We see that the fith area (from -200 to -80) is the dirtiest.

Unsurprisingly, it has the most of CD2 type debris

```
sum(data.1$cd2)
```

```
## [1] 2881345
```

```
sum(data.2$cd2)
```

```
## [1] 2638460
```

```
sum(data.3$cd2)
```

```
## [1] 0
```

```
sum(data.4$cd2)
```

```
## [1] 5434354
```

```
sum(data.5$cd2)
```

```
## [1] 15204656
```