Data Break 6: Meuse River Zinc, Empirical Semivariogram

BIOS 737 Spring 2004

Meuse River

- `library(gstat)`
- Data included in `gstat` library.
- `data(meuse)` makes the data available.
- Soil samples from flood plain of Meuse River, near village of Stein (Belgium).
- Geostatistical data: locations $\boldsymbol{s}$ and attributes $Z(\boldsymbol{s})$.

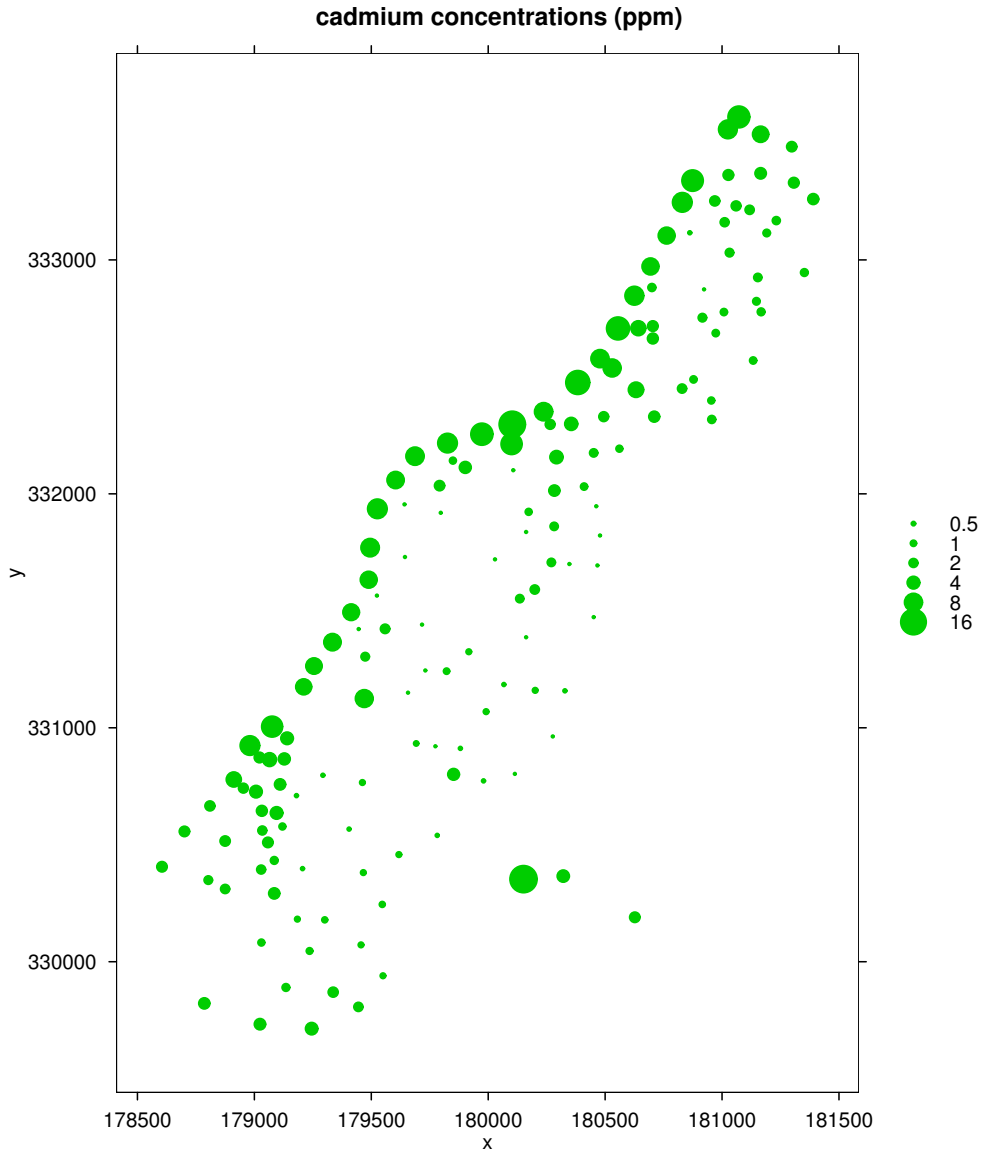
Visualizing data

- Create a “bubble plot” (dots at locations, size represents attribute).
- `bubble` command in `gstat` library.

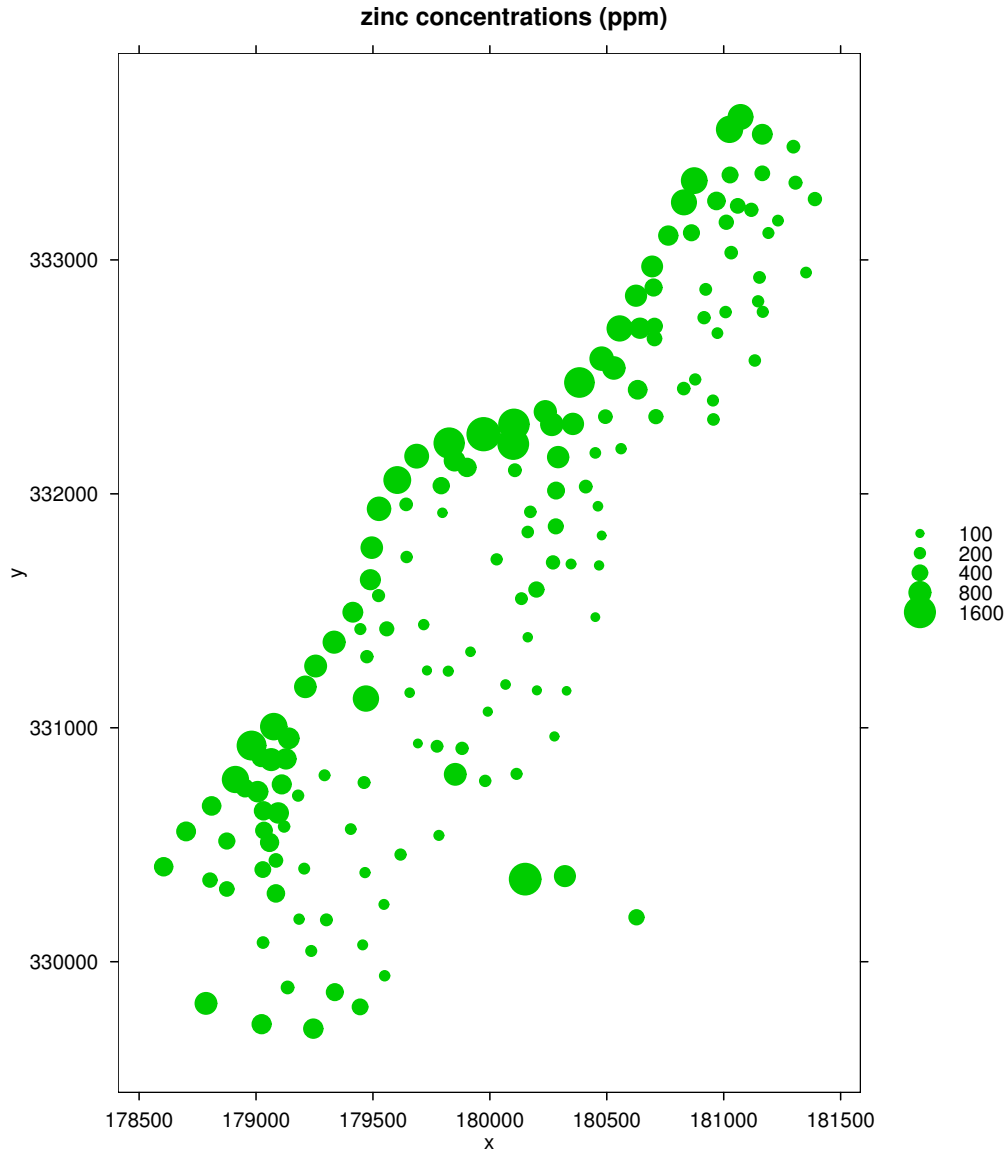
```
library(gstat)
data(meuse)

bubble(meuse, xcol="x",ycol="y",
    zcol="cadmium", maxsize = 2.5,
    main = "cadmium concentrations (ppm)",
    key.entries = c(.5,1,2,4,8,16))
```

Bubble plot: Cadmium



Bubble plot: Zinc

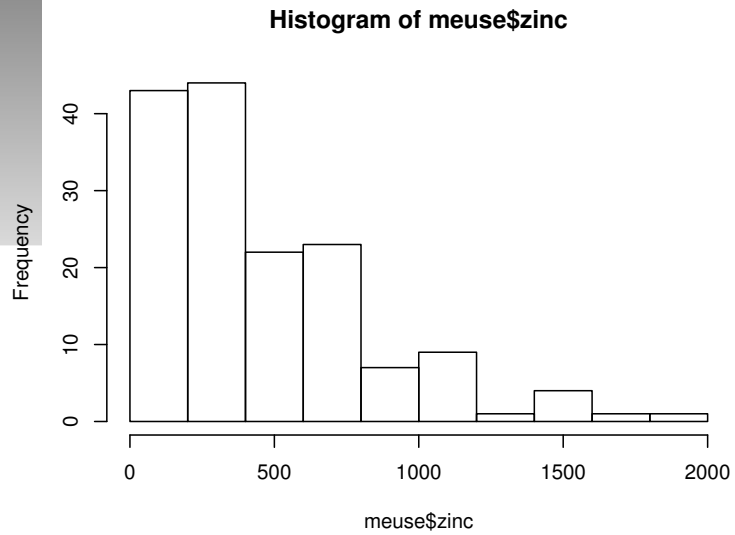


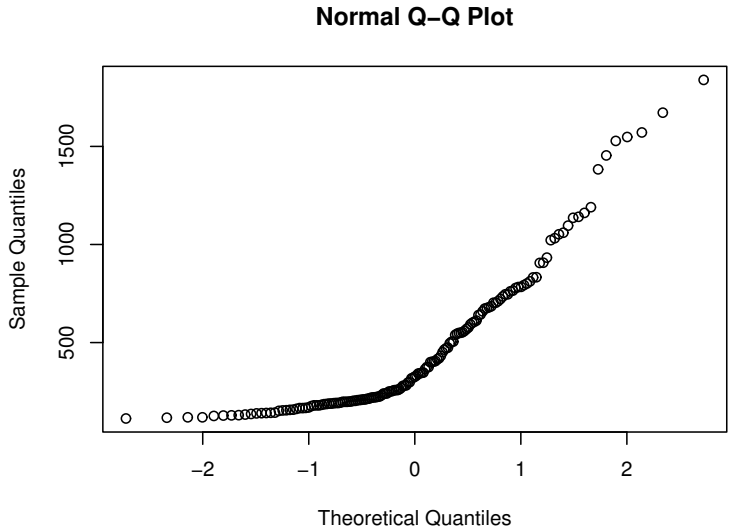
Our goal

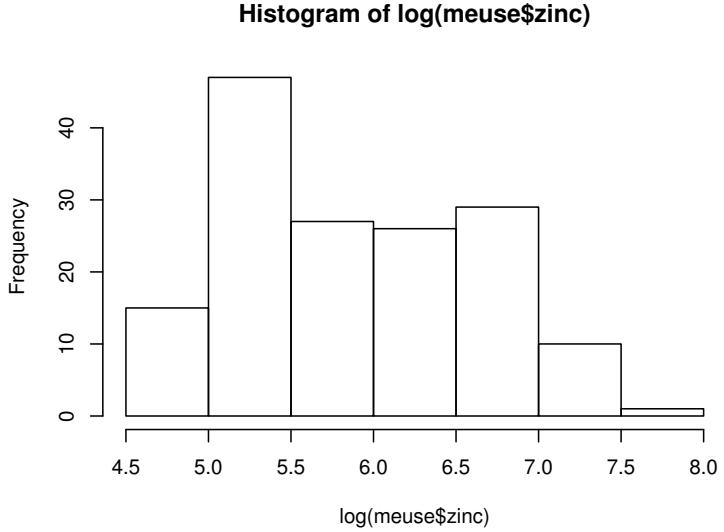
- Use (empirical) semivariogram to describe patter in zinc concentrations.
- Question 1: Do the zinc concentrations appear to be normally distributed?
- Log(zinc)?

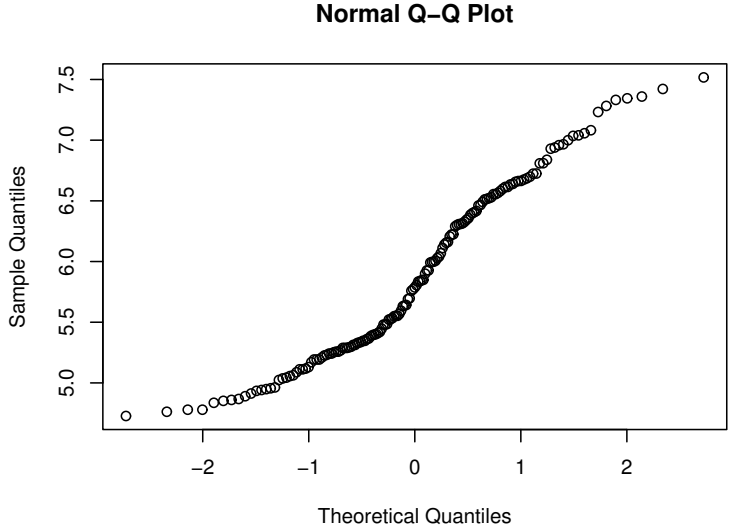
```
par(mfrow=c(2,2)
hist(meuse$zinc)
qqnorm(meuse$zinc)
hist(log(meuse$zinc))
qqnorm(log(meuse$zinc))
```

Normally distributed?









Using `gstat`

- Function `variogram` calculates the empirical variogram.
- Lots of options, we'll illustrate the most basic.
- `plot` recognizes output of `variogram` and plots with nice labels.

R *code*

```
logzinc.var <- variogram(log(zinc)~1,
                         loc=~x+y,
                         data=meuse)

print(logzinc.var)

plot(logzinc.var)
```

Output in `logzinc.var`

	np	dist	gamma	dir.hor	dir.ver
1	57	79.29244	0.1234479	0	0
2	299	163.97367	0.2162185	0	0
3	419	267.36483	0.3027859	0	0
4	457	372.73542	0.4121448	0	0
5	547	478.47670	0.4634128	0	0
6	533	585.34058	0.5646933	0	0
7	574	693.14526	0.5689683	0	0
8	564	796.18365	0.6186769	0	0
9	589	903.14650	0.6471479	0	0
10	543	1011.29177	0.6915705	0	0
11	500	1117.86235	0.7033984	0	0
12	477	1221.32810	0.6038770	0	0
13	452	1329.16407	0.6517158	0	0
14	457	1437.25620	0.5665318	0	0

Resulting plot