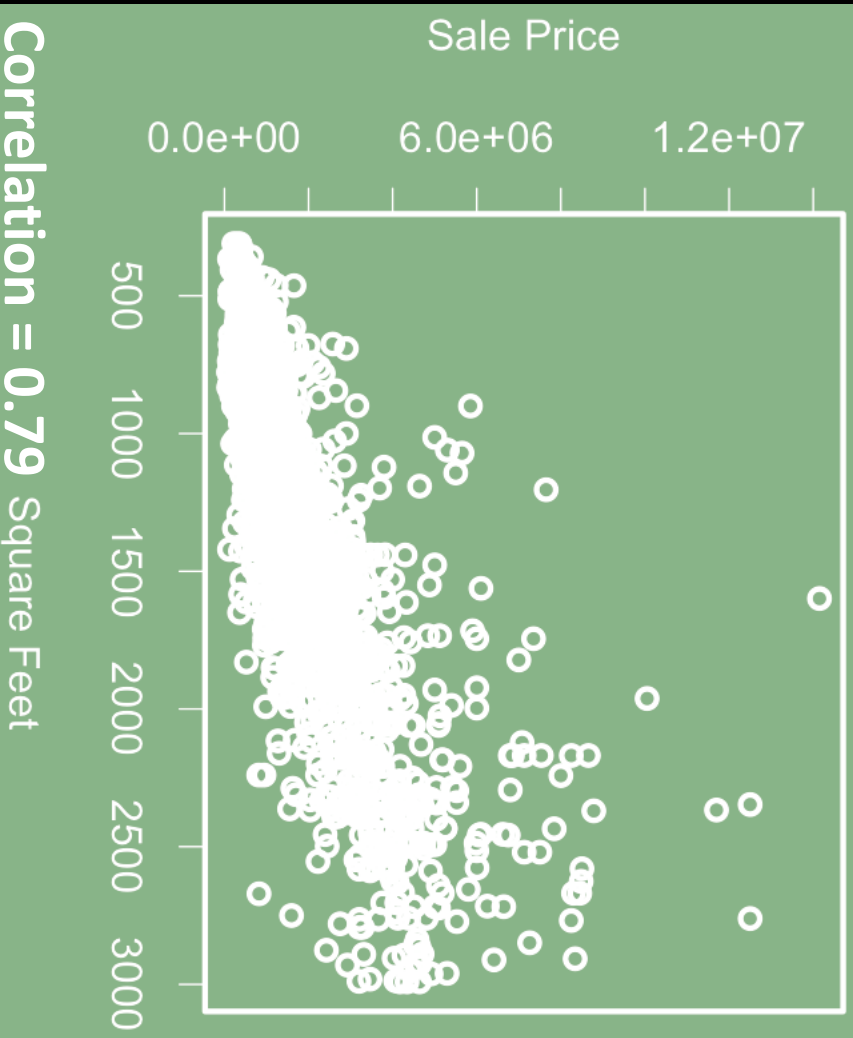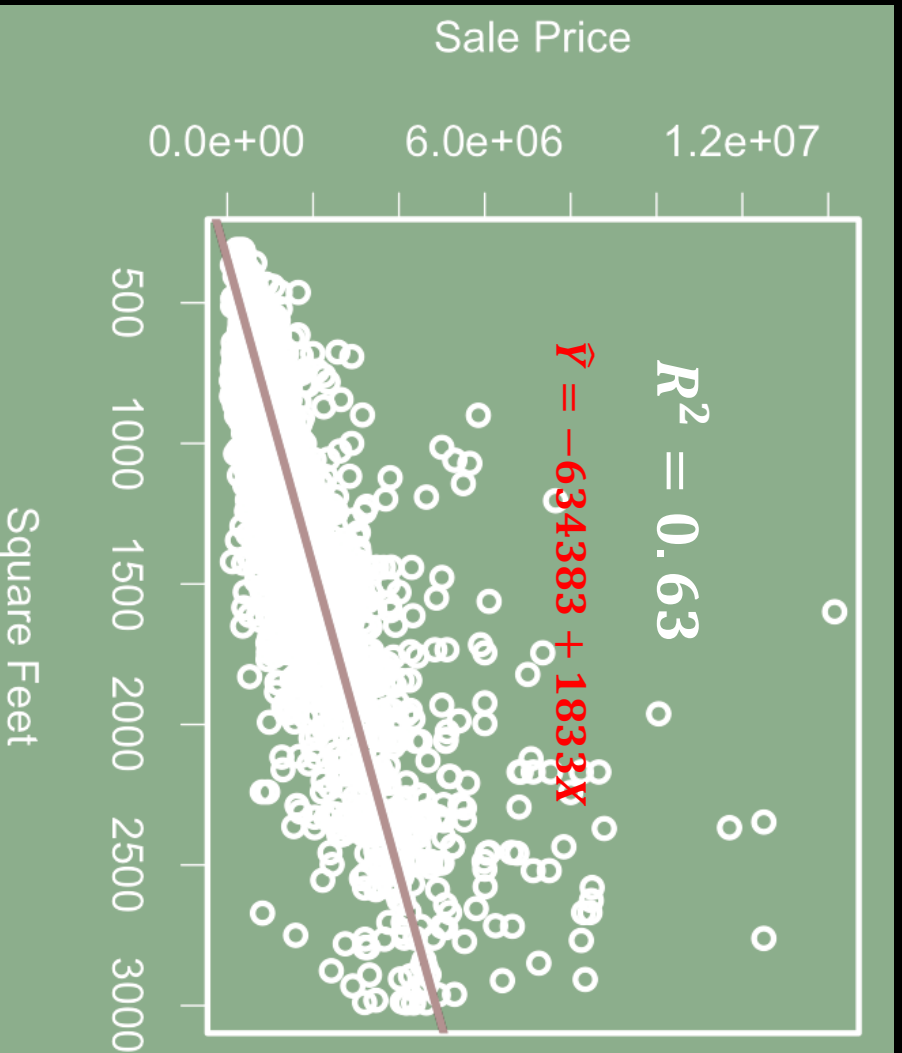# Manhattan Condo Prices

- From NYC Open Data
- Year 2009
- Condo building with elevator
- 3553 apartments
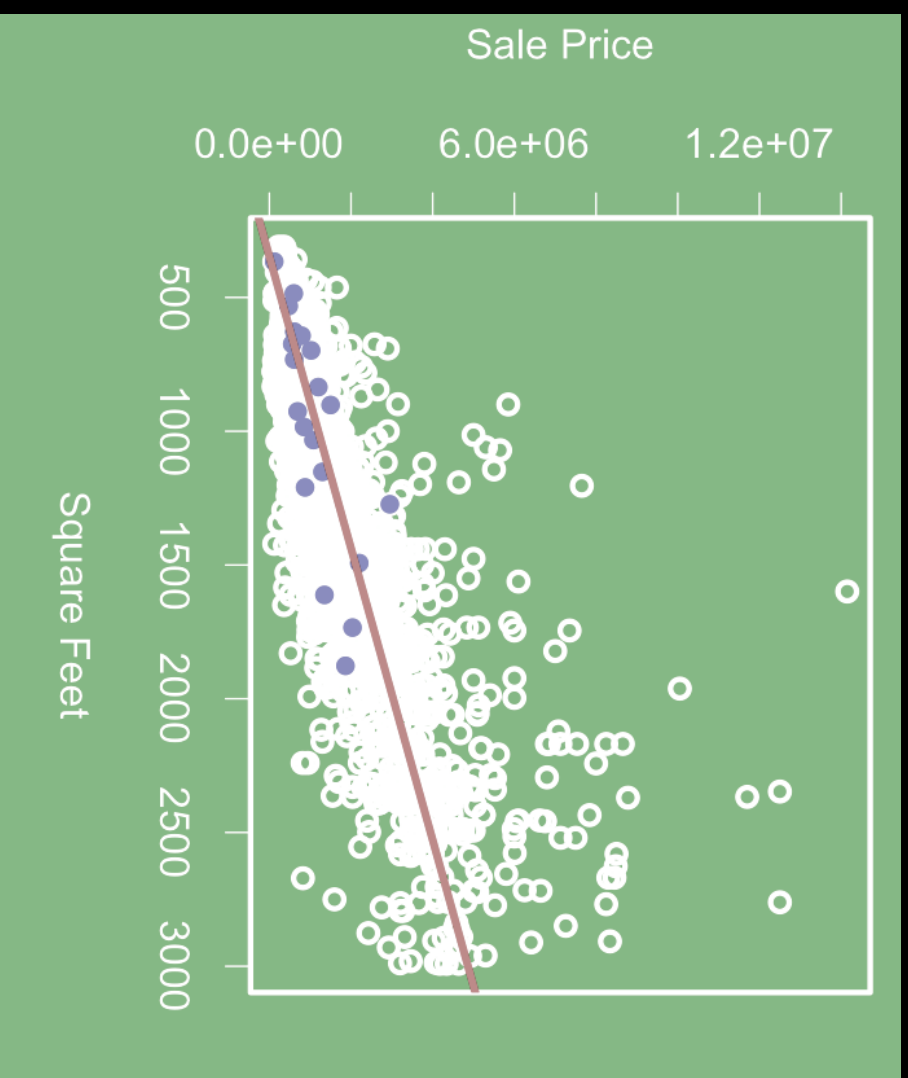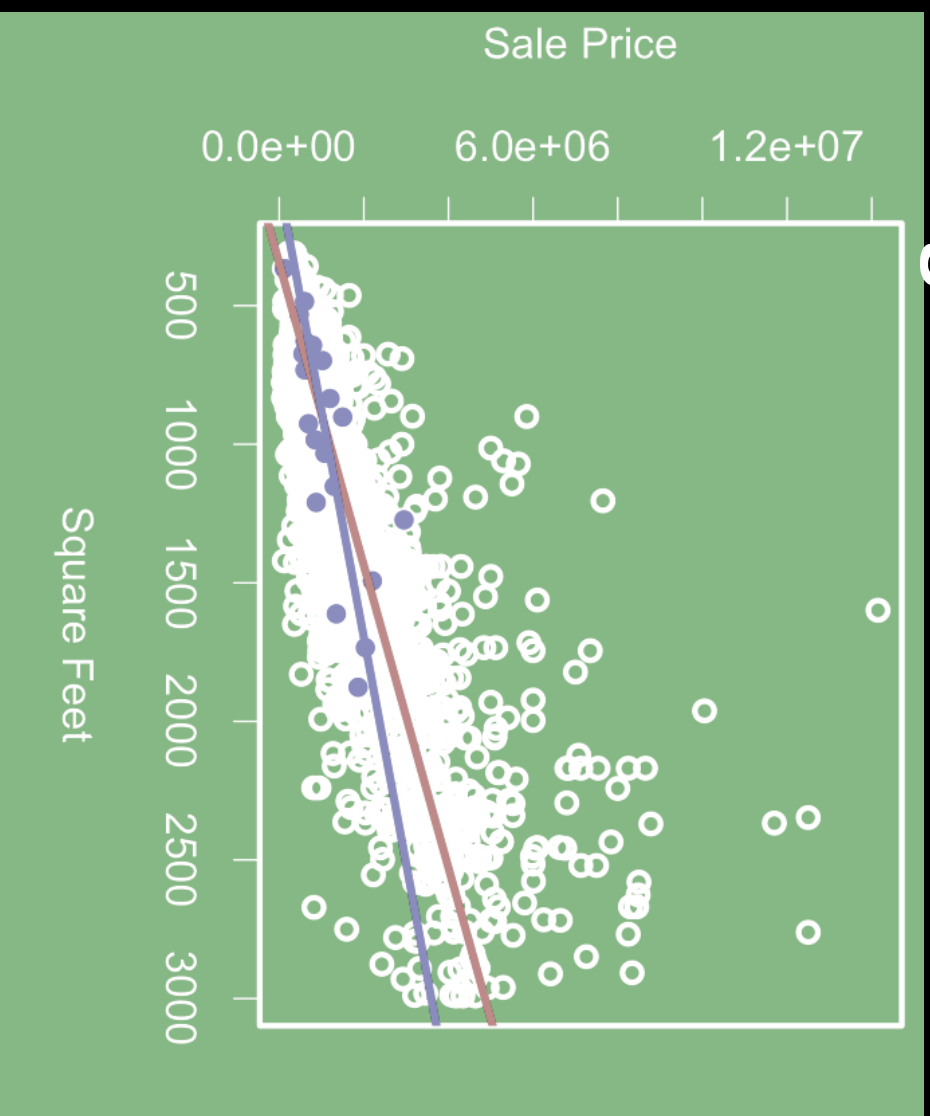
Manhattan Condo Prices

Manhattan Condo Prices

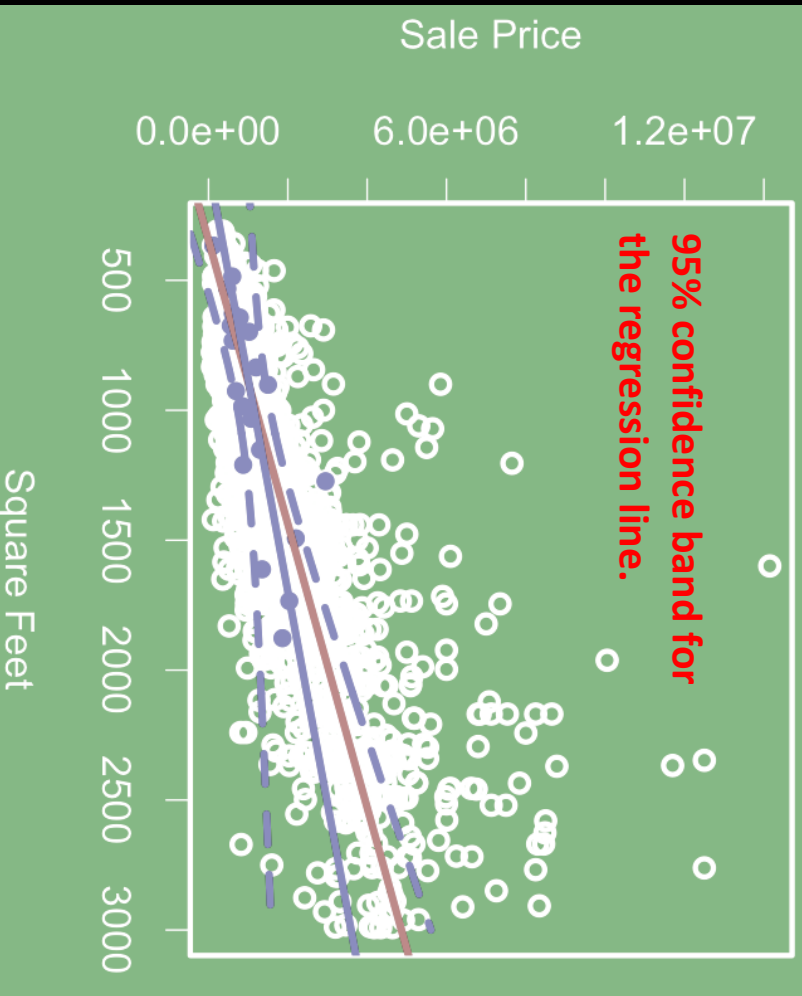Sampling variability in regression estimates

Sampling variability in regression estimates

Sampling variability in regression estimates

- **The confidence band centers at the sample estimate.**
- **It represents interval estimate for the regression line.**
- **Other inference on regression estimates can also be carried out.**

95% confidence band for the regression line.

Sale Price vs Square Feet

Prediction

- Given a value of X

- The predicted value is $\hat{Y} = b_0 + b_1 X$

- It is an estimate for the mean (average) value for Y given the X value.

- Most of the time, prediction is **different** from what is actually observed.
  - $Y - mean\ of\ Y$ (random variation)
  - $mean\ of\ Y - \hat{Y}$ (estimation error)

Prediction

- Extrapolation happens when one tries to give prediction on values of X outside the data range.

Multiple regression

- Y: response
- Multiple X variables
- $\hat{Y} = -546944 - 3265\,Age + 1770\,SQFT$
- Consider interaction
- $\hat{Y} = -743800 + 3137\,Age + 1996\,SQFT - 5.213\,Age{\times}SQFT$
- $\hat{Y} = (-743800 + 3137\,Age) + (1996 - 5.213\,Age)\,SQFT$

Multiple regression

Sale Price

Square Feet

0e+00   2e+06   4e+06   6e+06

500   1000   1500   2000   2500   3000

age=10
age=50
age=100