Concrete Plan Using Dr. Tang's Guidelines

Laughter segment
We are behind because of Elaine
1. Blame elaine and we get a curve
2. Make elaine fail so she doesn't graduate and loses her job
3. Steal Elaines future job everytime he gets it

Goal: Compare housing trends b/w the four most densely populated states
Ex: Avg sale to list strongest coeff. Of regression? But it may not be a great predictor for median sale price?
find interesting connections between diff columns
Actual meaning behind each potentially significant regressor

Project Outline
1. Create and analyze the fitted full model S.M., O.M.
2. Hypothesis Testing on Reg Coeffs. S.R.
   A. Test of Significance of Regression on a subset of regressors
   B. Analyze R^2 and adjusted R^2
3. Subset Selection S.M.
   A. Test on Subset of Regression Coeffs. (partial F-test)/GLH
   B. Include conf ints of mean response and prediction ints of parameters E.W.
4. Test for Multicollinearity (VIF) S.M.
5. Model Adequacy Checking (see Ch. 3.10 slide 13)
6. Residuals Analysis to identify points in question: outliers, leverage pts, and influential pts
   A. Influential Analysis (tests like DFBETAS, COVRATIO, etc.) O.M., E.W.
   B. Residual plots: normal prob plot and QQ plot E.W.
   C. Residuals vs. Fitted values and Residual O.M.
7. Transformations O.M., E.W.
8. Perform all possible regressions E.W.
9. Select models for further analysis (all)
10. Make recommendations and interpret regressors and results IN CONTEXT (all)

Sayema - will get started on the slides
- Analyze the full fitted model and create

Measure of Influence Organizer

Type of Measure	Which Statistics are Influenced	What it measures	Margin R uses vs. Margin taught in class/txt
Cook's D	Regression Coeffs	How much a point affects the conf int of the true reg coeff. Beta; i.e., the squared distance that the vec of fitted values moves when the ith obs is deleted	$D > F_{0.5,\,p,n-p}$ vs. $D > 1$
DFFITS	Single Fitted Value	Influence of deleting the ith obs on the prediction value	$\lvert DFFITS \rvert >$ $3 * \sqrt{p/(n-p)}$ vs. $\lvert DFFITS \rvert >$ $2 * \sqrt{p/n}$
DFBETAS	Regression Coeffs	How much a point affects the estimator of a specific (*jth*) reg coeff.	$\lvert DFBETAS \rvert > 1$ vs. $\lvert DFBETAS \rvert > 2/\sqrt{n}$
COVRATIO	Model Performance	Increase in precision of the regression line	COVRATIO $<$ $1 - 3 * p/(n-p)$ OR COVRATIO $>$ $1 + 3 * p/(n-p)$ vs. COVRATIO $<$ $1 - 3 * p/n$ OR COVRATIO $>$ $1 + 3 * p/n$
Hat Values	Variance of parameters: Model Accuracy	Indicates the potentially influential points	

Criteria for a reasonable transformation
1) Y seems to have a linear relationship with X: See R-squared and adjusted R-squared.
2) Enough scatter: uncorrelated errors
3) Constant variance, i.e., the scatter seems to be falling in between symmetrical bounds
4) Normally distributed residuals
5) Limited Outliers/influential points

Data Analysis
Data Description & Display (5)
S.M., E.W. Describes data and includes appropriate, well labeled, accurate graphs and tables of the data.

Data Analysis (10) Performs (i) model fitting, (ii) thorough residual analysis,
S.M., E.W. reasonable variable selection, and
O.M. (and E.W.) transformations if deemed necessary. Performs model fitting, residual analysis, and transformations if deemed necessary, but variable selection is not reasonable.

O.M. (with proofreading by everyone else) Conclusion (5) Includes a reasonable conclusion that is consistent with the data analysis. Reflective Process (5)