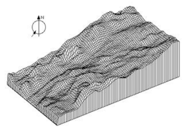
Module 8

Spatial Pattern: Smoothing and Trend Surface Analysis



Smoothing and Trend Surface

- **This module combines some previous ideas for a new application**
  - ***Search for spatial pattern (week 4):*** aids in helping us to understand the geographic structure of our datasets
  - ***Multiple regression (week 6):*** analysis that aids in building models to generalize our observations

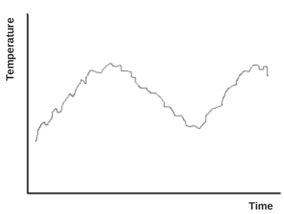
Smoothing and Trend Surface

- **A couple of key questions that we can address from this combined perspective**
  - **1. Can we "cut through the noise" to focus on the major trends in a map?**
    - Perhaps not important to represent every local variation in our model: major features only
  - **2. Can we build a model that represents the spatial patterns that we observe?**

Spatial Smoothing

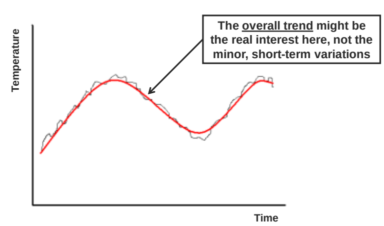
- **Smoothing can be thought of as simply "getting rid of the noise"**
- **A Simple Example: Smoothing a Time Series**
  - Although short-term variations can exist in a time series, they may be much less important than the longer-term trends reflected in the dataset

Time Series



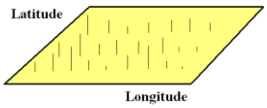
Original Data Series

Time Series



Smoothed Data Series

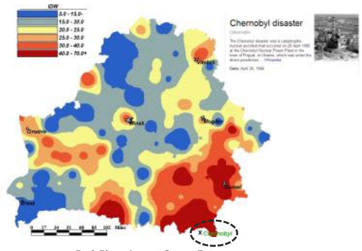
Geographic Data: A Parallel



Raw Data: A Digital Elevation Model

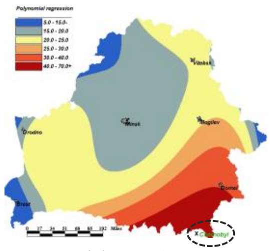
Idea would be to smooth this dataset so overall trends are apparent (not just individual point observations)

Lightly-Processed Map (Little Smoothing): Risk of Thyroid Cancer in Belarus



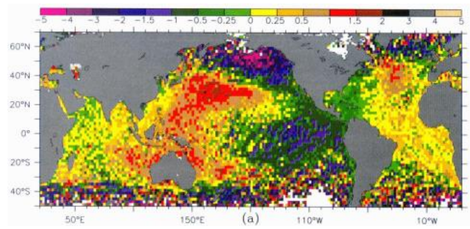
Dark Blue = Lowest Cancer Rate,
Dark Red = Highest Cancer Rate

Smoothed Map: Risk of Thyroid Cancer in Belarus



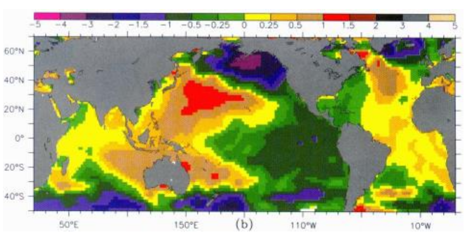
Dark Blue = Lowest Cancer Rate,
Dark Red = Highest Cancer Rate

Oceanic Air Pressures Example



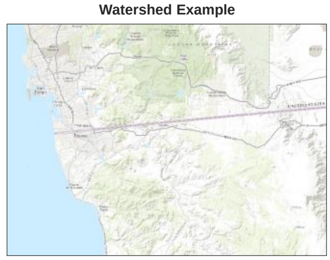
Source: NOAA, Pacific Marine Environmental Laboratory

The "Smoothed" Map



Source: NOAA, Pacific Marine Environmental Laboratory

Watershed Example



The US/Mexico Border at Tijuana/San Diego

Watershed Example

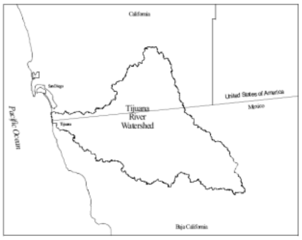


Figure 1 - The Tijuana River Watershed

Source: ESRI, "An ArcGIS Application of Spatial Statistics to Precipitation Modeling"

Watershed Example

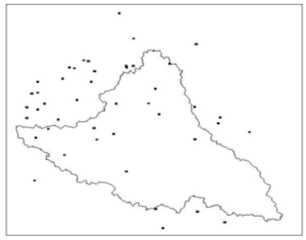


Figure 2 - Precipitation Stations for the Tijuana River Watershed

Source: ESRI, "An ArcGIS Application of Spatial Statistics to Precipitation Modeling"

Watershed Example

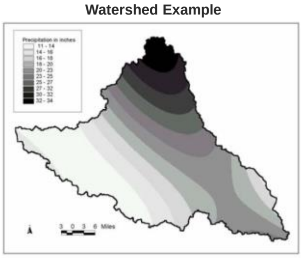


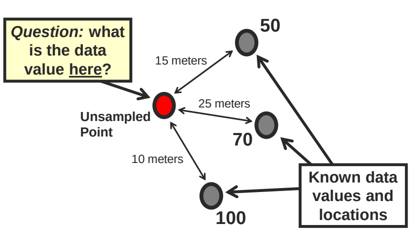
Figure 3 – Precipitation trend surface grid

Source: ESRI, "An ArcGIS Application of Spatial Statistics to Precipitation Modeling"

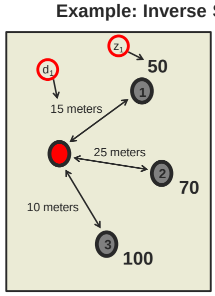
Spatial Smoothing

- **Basic Methods**
  - 1. Interpolation: "filling in the gaps"
    - Could simply do a visual estimate in a very simple situation (subjective)
    - *Better solution:* mathematical interpolation, such as using an "inverse squared model" (objective result)
    - Idea of mathematical interpolation is to calculate estimates based on data values at other nearby points

Example: Inverse Squared Interpolation



Example: Inverse Squared Interpolation



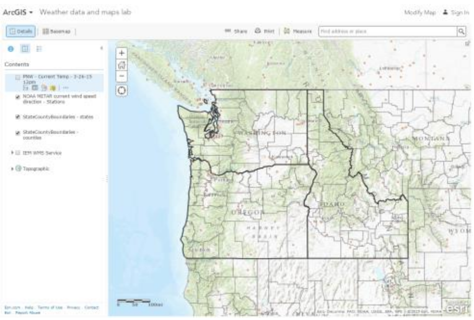
$$z = \frac{\left[\left(\frac{z_1}{d_1^2}\right)+\left(\frac{z_2}{d_2^2}\right)+\left(\frac{z_3}{d_3^2}\right)\right]}{\left[\left(\frac{1}{d_1^2}\right)+\left(\frac{1}{d_2^2}\right)+\left(\frac{1}{d_3^2}\right)\right]}$$

$$z = \frac{\left[\left(\frac{50}{15^2}\right)+\left(\frac{70}{25^2}\right)+\left(\frac{100}{10^2}\right)\right]}{\left[\left(\frac{1}{15^2}\right)+\left(\frac{1}{25^2}\right)+\left(\frac{1}{10^2}\right)\right]}$$

$$\boxed{z_{estimated} = 84.19}$$

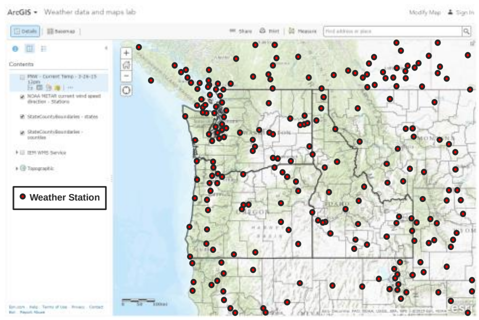
This is one objective interpolation method – there are many others

Weather Data Interpolation Example



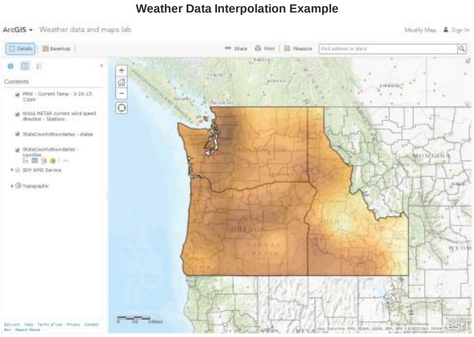
Courtesy of Brett Lucas

Weather Data Interpolation Example



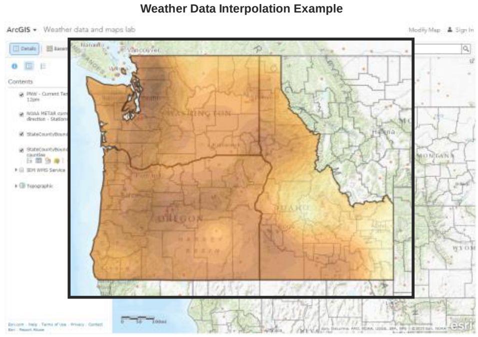
Courtesy of Brett Lucas

Weather Data Interpolation Example



Courtesy of Brett Lucas

Weather Data Interpolation Example



Courtesy of Brett Lucas

Spatial Smoothing

- **Basic Methods**
  - 2. Running Mean: "a local average"
    - Can think of running means in both 1- and 2-dimensional cases
    - ***1-dimensional:*** as in a time series, or a series of observations taken every 100 feet along a highway (a series of numbers)
    - ***2-dimensional:*** as in a GIS data file with x and y coordinates (spatial attributes)

Example: Running Mean in One Dimension

Original Data Series	10	12	15	20
Average of 2	11	13.5	17.5	
Average of 3	12.3	15.7		

Example: Running Mean in Two Dimensions

"Two Dimensional Filter Mapping"

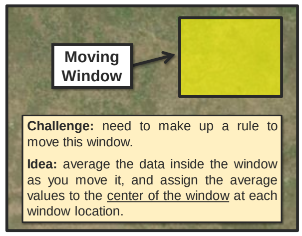
Study Area

Each pixel in the satellite image has a value associated with it.

Example: Running Mean in Two Dimensions

"Two Dimensional Filter Mapping"

Study Area



Example: Running Mean in Two Dimensions

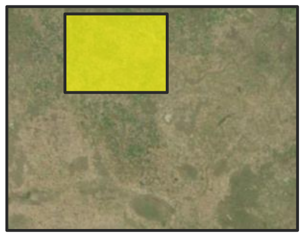
"Two Dimensional Filter Mapping"

Study Area

Example: Running Mean in Two Dimensions

"Two Dimensional Filter Mapping"

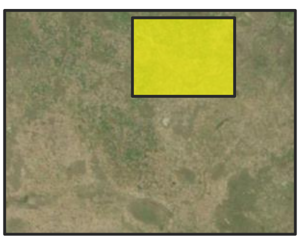
Study Area



Example: Running Mean in Two Dimensions

"Two Dimensional Filter Mapping"

Study Area



Example: Running Mean in Two Dimensions

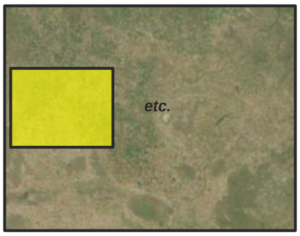
"Two Dimensional Filter Mapping"

Study Area

Example: Running Mean in Two Dimensions

"Two Dimensional Filter Mapping"

Study Area

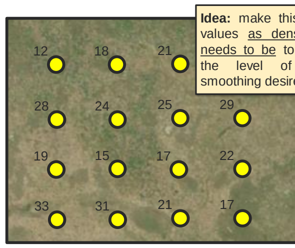


Example: Running Mean in Two Dimensions

"Two Dimensional Filter Mapping"

Study Area

Assign <u>average values to each window center location</u> as we move the window across the study area



Idea: make this grid of values <u>as dense as it needs to be</u> to achieve the level of spatial smoothing desired.

Spatial Smoothing

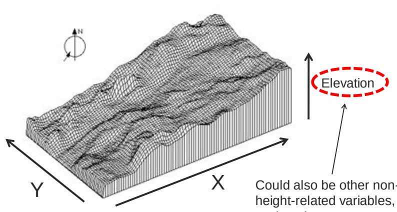
- **Problems with two dimensional filter mapping**
  - 1. how do you <u>move</u> the window?
  - 2. what window <u>size</u> and <u>shape</u> do you use?
  - 3. what do you do at the <u>edges</u> of your study area? (we'll see a map in a few minutes that illustrates this issue further)

Spatial Smoothing

- **Spatial smoothing is a very useful analytical method**
  - But for the results to be meaningful for your study, you must think through your options

Trend Surface Analysis

- **Basic Concepts**
  - TSA is a technique for combining <u>modeling</u> with <u>smoothing</u>
  - TSA is really just <u>multiple regression</u> with location coordinates (longitude, latitude) as the independent variables
  - The variable of interest for your research (elevation, temperature, income, average age, etc.) is the dependent variable



Could also be other non-height-related variables, such as income, population density, moisture level, etc.

Trend Surface Analysis

- **Basic Concepts**
  - Typically start with a linear, or first order, model (a plane)

    $$Elevation = a + b_1x + b_2y$$

  - Could next consider a second order model (a quadratic surface) if needed

    $$Elevation = a + b_1x + b_2y + b_3x^2 + b_4y^2 + b_5xy$$

  - Could continue to higher order models (cubic models and on, only if necessary)

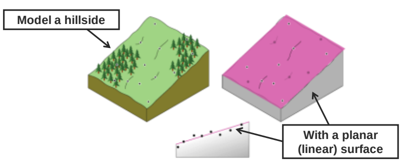
Breaking this Quadratic Expression into Components…

$$Elevation = a + \underbrace{\boxed{b_1x}}_{\text{Linear term in } x} + \underbrace{\boxed{b_2y}}_{\text{Linear term in } y} + \underbrace{\boxed{b_3x^2}}_{\text{Quadratic term in } x} + \underbrace{\boxed{b_4y^2}}_{\text{Quadratic term in } y} + \underbrace{\boxed{b_5xy}}_{\text{Term representing the interaction between } x \text{ and } y}$$

So what does this trend surface modeling look like in real-world terms?

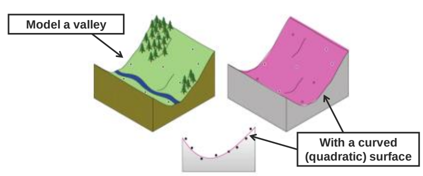
An example of straightforward elevation modeling with a planar surface:

An example of straightforward elevation modeling with a planar surface:



Source: ESRI, Online ArcGIS Documentation

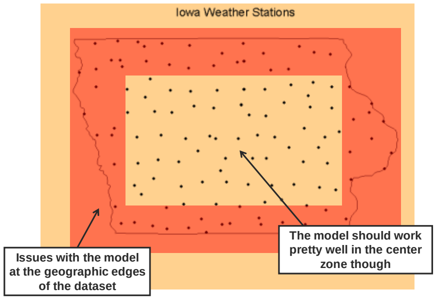
Another elevation modeling example with a quadratic surface:



Source: ESRI, Online ArcGIS Documentation

Trend Surface Analysis

- **Basic Concepts**
  - Straightforward technique and application, but several potential problems, particularly with higher order models (*i.e.* cubic and higher equations)
    - Surfaces can become "unstable" at map edges: the edge effect



Source: UCSB, Department of Geography

Trend Surface Analysis

- **Basic Concepts**
  - Other complications
    - Watch out for numerical problems, particularly with some coordinate systems (*e.g.* UTM)
      - **Important point:** make sure you don't mix coordinate systems within a single model (one variable mapped in one coordinate system, a second variable in a different coordinate system)

Trend Surface Analysis

- **Basic Concepts**
  - Other complications
    - Residuals are likely to show <u>spatial autocorrelation</u>
    - Remember what we said earlier in the course: <u>this is usually not a "problem"</u> for geographers

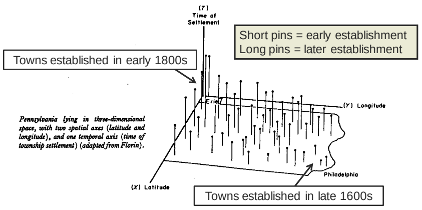
Trend Surface Analysis

- **Basic Concepts**
  - What's "exciting" about studying spatial autocorrelation with residuals in TSA?
    - When we uncover spatial autocorrelation we are really finding some <u>clues</u> that can help us build even better spatial models
    - Essentially, spatial autocorrelation is often an input to <u>new theory generation</u> (a good thing)

TSA and Residual Analysis

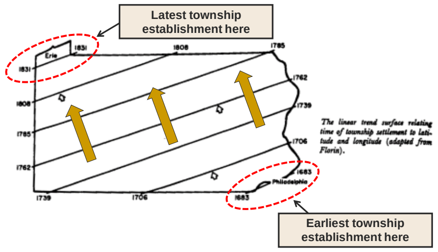
- **Let's use an example to see how this <u>residual analysis</u> can really help**
  - ***Case study:*** trend surface analysis of settlement development in Pennsylvania
    - <u>Goal</u>: generate a spatial model of *when towns and cities were established* across the state
    - This entire analysis (basic TSA model + follow-up examination of residuals) can provide real insight into the <u>processes</u> that guided city establishment in PA

Pennsylvania Settlement Example: Raw Data



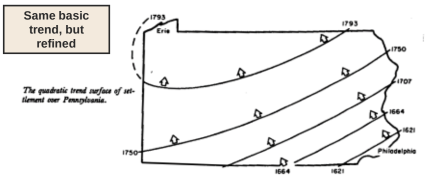
Source: Abler, Ronald F., John S. Adams, and Peter Gould (1971) *Spatial Organization: The Geographer's View of the World*. Englewood Cliffs, NJ: Prentice-Hall, page 135.

Pennsylvania Settlement Example: Linear Trend Surface



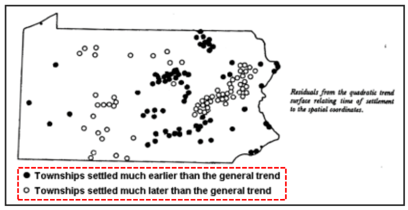
Source: Abler, Ronald F., John S. Adams, and Peter Gould (1971) *Spatial Organization: The Geographer's View of the World*. Englewood Cliffs, NJ: Prentice-Hall, page 135.

Pennsylvania Settlement Example: Quadratic Trend Surface



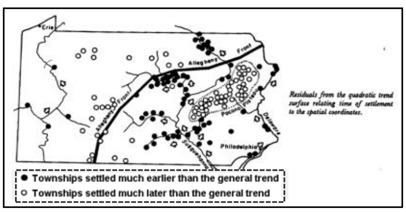
Source: Abler, Ronald F., John S. Adams, and Peter Gould (1971) *Spatial Organization: The Geographer's View of the World*. Englewood Cliffs, NJ: Prentice-Hall, page 135.

Pennsylvania Settlement Example: Residual Analysis



How could we interpret this residual pattern?

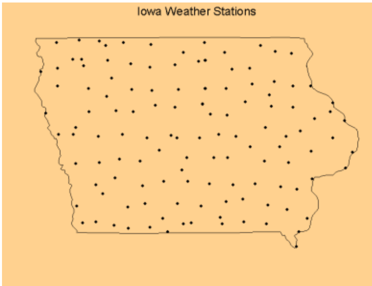
Pennsylvania Settlement Example: Residual Analysis



How could we interpret this residual pattern?

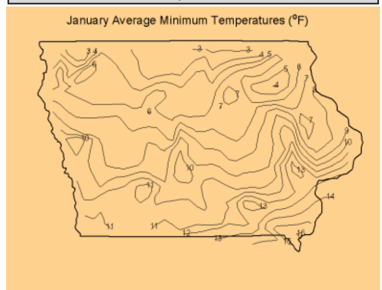
Another TSA Example

- Let's follow through another, complete "Iowa weather station" case study to see some of the <u>analytical details</u> and <u>thought processes</u> related to how we can actually apply the TSA technique
  - The following gives some idea of what is possible in <u>refining</u> your trend surface model, starting with your basic dataset



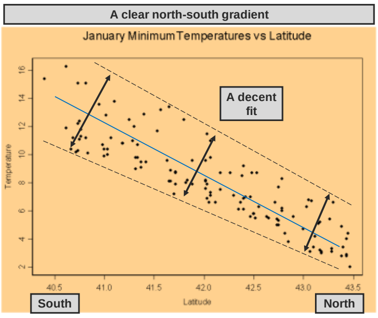
Source: UCSB, Department of Geography

Contour map of weather data



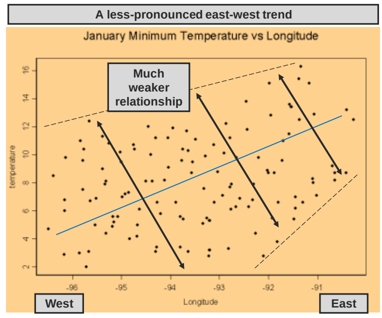
Source: UCSB, Department of Geography

A clear north-south gradient



Source: UCSB, Department of Geography

A less-pronounced east-west trend



Source: UCSB, Department of Geography

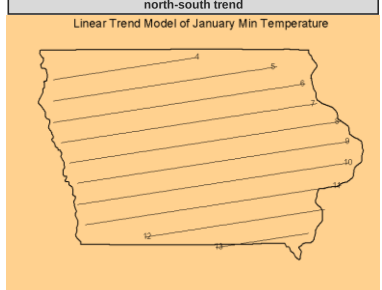
First step: try fitting a linear trend surface to the weather dataset

In other words, fit a TSA equation of the form:

$$Temperature = a + b_1 \times Latitude + b_2 \times Longitude$$

This is just a basic multiple regression analysis with
- **<u>One dependent variable</u> (temperature) and**
- **<u>Two independent variables</u> (latitude and longitude)**

Planar-fitted surface reflects the clear and dominant north-south trend



Source: UCSB, Department of Geography

Linear Model: Summary of Results

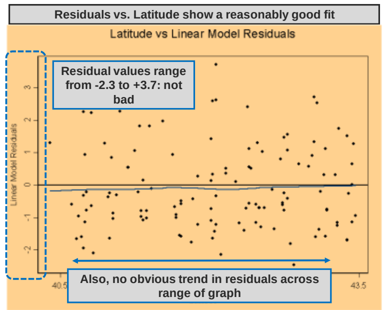
```
Coefficients:
                Value Std. Error  t value Pr(>|t|)
(Intercept)    8.1545   0.1236    65.9728   0.0000
        lta   -3.1318   0.1447   -21.6452   0.0000
        lna    0.3587   0.0766     4.6806   0.0000

Residual standard error: 1.371 on 120 degrees of freedom
Multiple R-Squared: 0.8162
F-statistic: 266.4 on 2 and 120 degrees of freedom, the p-value is 0
```

Highly significant

High R^2

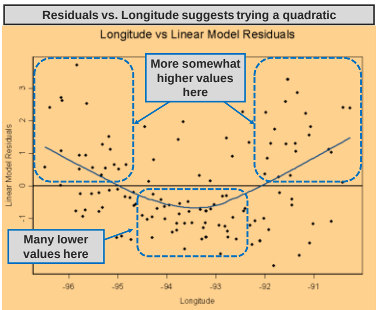
Results show a <u>fairly high R-squared</u> for a model that is <u>statistically significant</u> (so the trend surface model fits the weather data pretty well)

Residuals vs. Latitude show a reasonably good fit



Source: UCSB, Department of Geography

Residuals vs. Longitude suggests trying a quadratic



Source: UCSB, Department of Geography

Next step: try fitting a quadratic surface, as suggested by the residual analysis

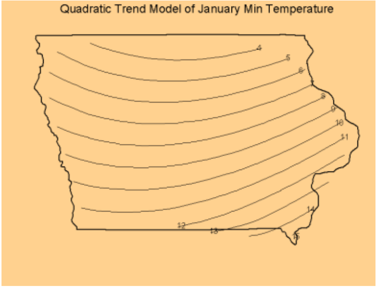
In other words, fit a TSA equation of the form:

$$Temperature = \mathrm{a} + b_1 \times lat + b_2 \times long + b_3 \times long^2 + b_4 \times lat \times long$$

In this case, and as suggested by the residual analysis:

- <u>Linear</u> with latitude and longitude
- <u>Quadratic</u> with longitude
- Also trying an <u>interaction</u> term (latitude and longitude jointly)

Fitted surface now quadratic in east-west direction



Source: UCSB, Department of Geography

Quadratic Model: Summary of Results

Including <u>squared longitude</u> and <u>longitude times latitude</u>

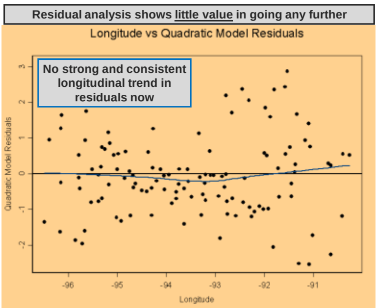
```
Coefficients:
                  Value Std. Error  t value Pr(>|t|)
  (Intercept)    7.5050   0.1557    48.2040   0.0000
          lta   -3.2493   0.1268   -25.6327   0.0000
          lna    0.3468   0.0670     5.1751   0.0000
    I(lna^2)     0.2272   0.0435     5.2242   0.0000
I(lta * lna)    -0.2130   0.0845    -2.5210   0.0130

Residual standard error: 1.188 on 118 degrees of freedom
Multiple R-Squared: 0.8643
F-statistic: 187.9 on 4 and 118 degrees of freedom, the p-value is 0
```

Still highly significant

Even Higher R^2

Results show a <u>fairly substantial improvement</u> (in R^2) on the already-good initial model, and still <u>highly statistically significant</u>

Residual analysis shows little value in going any further



Source: UCSB, Department of Geography

TSA Conclusion

- Note that you don't necessarily have to have access to a GIS to do a TSA
  - TSA is basically multiple regression, so SPSS alone can get you started with TSA model building
  - However, using a GIS for TSA has the advantage of
    - letting you see the results of your TSA on a map (as with the Iowa and Pennsylvania examples)
    - gain a spatial perspective on your residuals (Pennsylvania example)