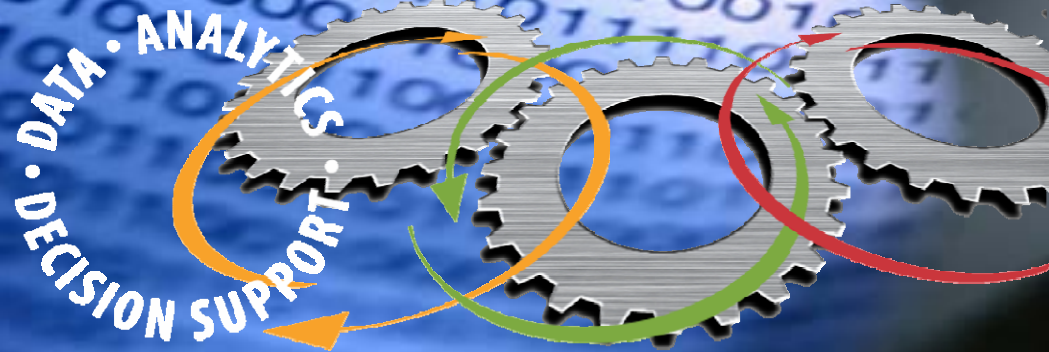




# Practical Data Mining & Modeling

Jennifer Lemus, ISO Innovative Analytics



# Overview

- **Data gathering and data management**
- **Exploration**
- **Model building**
- **Testing and validation**

# Data Gathering and Data Management

- 60% + of effort
- Need a data expert
- Real world data is messy
- Analysts need exposure to the 'raw' data

# Exploration

- **This stage is often an iterative process with the data management effort**
- **Requires subject matter knowledge to wade through the chaff**
- **Also requires an unbiased eye so nothing is overlooked**

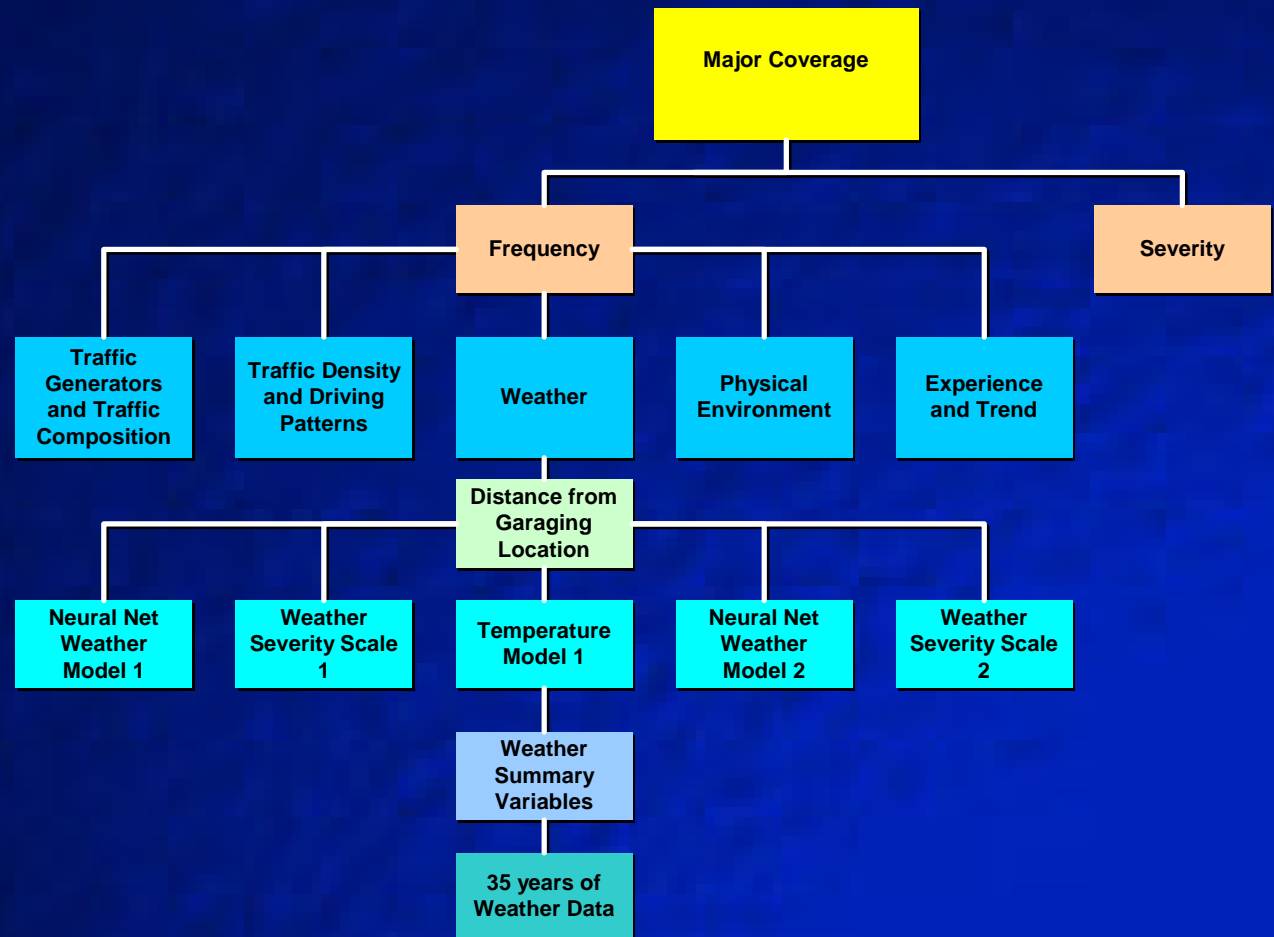
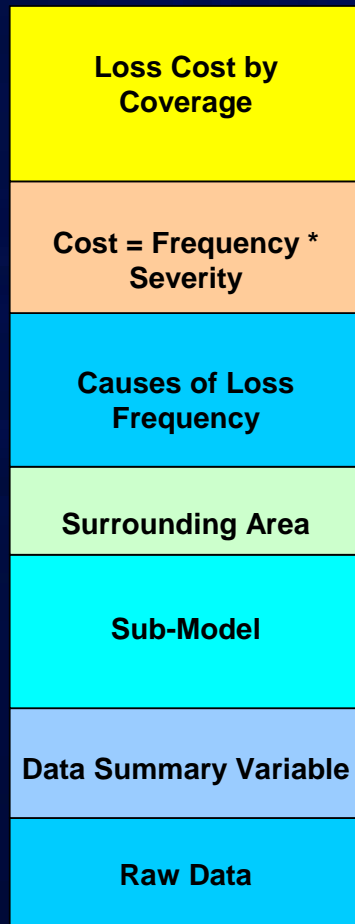
# Exploratory Techniques

- **Decision trees**
- **Tables**
- **Univariate statistics**
- **Dimension reduction techniques**
- **Exploration drives the direction of the modeling**

# Model Building

- Use exploratory tools to build sub models
- Use model or models as inputs into rules based system – the model is not necessarily the final product

# Using Weather to Predict Insurance Loss



# Testing and Validation

- **Test fit**
- **Analyze residuals**
- **Common sense checks**
- **Watch out for tautologies**
- **Test cases**

# Conclusion

"All models are wrong but some are useful"

George E.P. Box