

# Learning Objectives for Data Concept and Visualization

## Assignment 1: Data Quality

MODULE TITLE	LEARNING OBJECTIVES
<b>Concept and Impact of Data Quality</b>	<ul style="list-style-type: none"> <li>Summarize concepts of data quality. Understand and describe the impact of data on actuarial work and projects.</li> </ul>
<b>Data Quality Principles</b>	<ul style="list-style-type: none"> <li>Understand the categories of data quality principles. Given a principle of data quality, provide an example that illustrates the principle. Understand what is involved in a review of data.</li> </ul>
<b>Data Governance</b>	<ul style="list-style-type: none"> <li>Concepts, roles and responsibilities, committees, tools</li> </ul>
<b>Data Documentation: Metadata Terminology</b>	<ul style="list-style-type: none"> <li>Describe these aspects of data documentation:               <ul style="list-style-type: none"> <li>➤ Data and metadata terminology</li> <li>➤ Relationship between data documentation and data governance</li> <li>➤ Types and uses of metadata for data scientists</li> </ul> </li> </ul>
<b>Aggregate Insurance and Statistical Data</b>	<ul style="list-style-type: none"> <li>Explain the regulator and business needs for statistical data</li> </ul>

## Assignment 2: Sources of Data

MODULE TITLE	LEARNING OBJECTIVES
<b>Data Science and Data Scientists</b>	<ul style="list-style-type: none"> <li>Understand the fundamental concepts of data science.</li> <li>Understand different types of data scientists.</li> <li>Summarize the insurers' use of predictive analytics and data science and the roles of data scientists and data science team.</li> </ul>
<b>Insurer Data-Driven Decision Making</b>	Explain how insurers and risk managers use data-driven decision making
<b>Insurer Operational Data</b>	<ul style="list-style-type: none"> <li>Understand what typical attributes are made available in each of the following data sources:               <ul style="list-style-type: none"> <li>➤ Policy and Premium Data</li> <li>➤ Claims Information</li> <li>➤ Claim Notes</li> <li>➤ Billing Information</li> <li>➤ Producer Information</li> </ul> </li> <li>Understand how corrections for each of these attributes are recorded.</li> </ul>

	<ul style="list-style-type: none"> <li>• Understand timing of collection and updating. Understand how quality can change over time.</li> </ul>
<b>The Value of Statistical Plan Data</b>	<ul style="list-style-type: none"> <li>• Understand why insurance companies produce statistical files, typical attributes in stat files and the advantages and disadvantages of using stat files over operational data.</li> </ul>
<b>Statistical Plans</b>	<ul style="list-style-type: none"> <li>• Describe the two basic types of statistical plans and their use in insurers' predictive modeling and ratemaking</li> </ul>
<b>Insurance Date Fields and Amount Fields in Statistical Plans</b>	<ul style="list-style-type: none"> <li>• Explain the role of the statistical agent and how the date and amount fields are used in the statistical plans.</li> </ul>
<b>Classification or Rating Variable Fields and Exposure Data Elements in Statistical Plans</b>	<ul style="list-style-type: none"> <li>• Describe the following statistical plan elements by line of business. <ul style="list-style-type: none"> <li>➤ Classification and rating variable fields</li> <li>➤ Exposure</li> </ul> </li> </ul>
<b>External Sources of Noninsurance Data</b>	<ul style="list-style-type: none"> <li>• Understand how to access and the uses of external sources of data including: <ul style="list-style-type: none"> <li>➤ Demographic information</li> <li>➤ Customer Financial Information</li> <li>➤ Business Financial Information</li> <li>➤ Behavioral data</li> <li>➤ Driving Records and Motor Vehicle Reports</li> <li>➤ Government sources</li> </ul> </li> <li>• Understand who collects the information, for what purposes, how frequently it is updated and how it is distributed. For each of these sources understand what typical attributes are made available.</li> <li>• Understand various derived attributes. Understand if there is a clear way to merge the data into databases used for analysis.</li> </ul>
<b>External Sources of Data – Insurance Specific</b>	<ul style="list-style-type: none"> <li>• Understand External sources of Insurance Specific data such as: <ul style="list-style-type: none"> <li>➤ Historical claims reports</li> <li>➤ Industry trend factors</li> <li>➤ Loss Development factors</li> <li>➤ Data available in NCCI Statistical Bulletin</li> </ul> </li> </ul>
<b>External Data Sources: United States Census Data</b>	<ul style="list-style-type: none"> <li>• Understand the two basic formats into which data from the United States Census bureau is organized.</li> </ul>

### Assignment 3: Accessing Raw Data

MODULE TITLE	LEARNING OBJECTIVES
<b>Data Classifications</b>	<ul style="list-style-type: none"> <li>• Understand the broad classifications of data: quantitative vs qualitative; nominal, ordinal, interval continuous; transactional, snapshots, aggregated.</li> <li>• Understand transaction and snapshot data and how to combine snapshots from various times into transaction files.</li> </ul>
<b>Working with Structured Data</b>	<ul style="list-style-type: none"> <li>• Understand the various data types: Numeric, string, date, geographic.</li> <li>• Understand how to read, store and display each data type. Understand issues with Date types, such as different formatting across different data sources-coding and why it is used.</li> <li>• Understand structured versus unstructured data and the various forms of unstructured data such as document, map, voice, and image.</li> </ul>
<b>Unicode Basics</b>	<ul style="list-style-type: none"> <li>• Explain the purpose and functions of Unicode.</li> </ul>
<b>Working with Unstructured Data</b>	<ul style="list-style-type: none"> <li>• Be aware of other external sources of data such as:               <ul style="list-style-type: none"> <li>➤ Social media feeds</li> <li>➤ Web APIs</li> <li>➤ ACID</li> <li>➤ NoSQL</li> </ul> </li> </ul>
<b>Text File Formats</b>	<ul style="list-style-type: none"> <li>• Explain how to read and write data to delimited and fixed text file formats</li> </ul>
<b>Dataframes</b>	<ul style="list-style-type: none"> <li>• Explain how data can be entered and stored in these types of dataframes.               <ul style="list-style-type: none"> <li>➤ Relational databases</li> <li>➤ Data warehouses</li> <li>➤ Excel</li> <li>➤ R</li> <li>➤ Python</li> </ul> </li> </ul>
<b>Data Exchange</b>	<ul style="list-style-type: none"> <li>• Describe how the exchange of data is facilitated by:               <ul style="list-style-type: none"> <li>➤ HTML</li> <li>➤ XML</li> <li>➤ JSON</li> <li>➤ Data marts</li> </ul> </li> </ul>
<b>Obtaining Data from the Internet</b>	<ul style="list-style-type: none"> <li>• Describe methods for obtaining data from the internet</li> </ul>

<b>Data Profiling</b>	<ul style="list-style-type: none"> <li>• Demonstrate ability to profile data including: inspecting rows of data, read data catalogs and metadata. Create descriptive statistics and graphs that profile the data.</li> </ul>
<b>Messy Data</b>	<ul style="list-style-type: none"> <li>• Detect and remediate missing, miscoded and anomalous data. Understand sampling bias and clustering of values.</li> </ul>
<b>Testing Data</b>	<ul style="list-style-type: none"> <li>• Work with small test data and create sample data using simple filters. Be able to sample from related tables.</li> </ul>

**Assignment 4: Working with Data**

<b>MODULE TITLE</b>	<b>LEARNING OBJECTIVES</b>
<b>Querying Data</b>	<ul style="list-style-type: none"> <li>• Explain how to query data from a database using Structured Query Language <ul style="list-style-type: none"> <li>➤ Querying Data</li> <li>➤ SELECT, FROM, and WHERE Statements</li> <li>➤ Retrieving Columns</li> <li>➤ Retrieving Distinct Data</li> <li>➤ Aggregate Functions</li> <li>➤ Grouping Data</li> </ul> </li> </ul>
<b>Joining Data Tables</b>	<ul style="list-style-type: none"> <li>• Explain how to join data tables using Structured Query Language <ul style="list-style-type: none"> <li>➤ Multiple Tables</li> <li>➤ Joining Two Tables</li> <li>➤ Joining More Than Two Tables</li> <li>➤ Subqueries</li> </ul> </li> </ul>
<b>Advanced SQL Topics</b>	<ul style="list-style-type: none"> <li>• Describe these issues in Structured Query Language: Indexes; Null values; User-defined functions; Large-data access <ul style="list-style-type: none"> <li>➤ Indexes</li> <li>➤ Null Values</li> <li>➤ User-Defined Functions</li> <li>➤ Large-Data Access</li> </ul> </li> </ul>
<b>String Functions</b>	<ul style="list-style-type: none"> <li>• Understand the use of string functions in common data preprocessing software. <ul style="list-style-type: none"> <li>➤ String Processing With SQL</li> <li>➤ String Processing With Regular Expressions</li> </ul> </li> </ul>

<b>Working With Regular Expressions in String Processing</b>	<ul style="list-style-type: none"> <li>• Understand the use of common regular expressions for pattern matching. <ul style="list-style-type: none"> <li>➤ Basic Regular Expressions</li> <li>➤ Application of Regular Expressions</li> </ul> </li> </ul>
<b>Using Hash Tag Functions</b>	<ul style="list-style-type: none"> <li>• Explain how to use hash functions with databases. <ul style="list-style-type: none"> <li>➤ Common Hash Functions</li> <li>➤ Hash Functions for Equality Testing</li> <li>➤ Hash Functions for Hash Tables</li> <li>➤ Hash Functions for Data Segmentation</li> </ul> </li> </ul>
<b>Insurance Applications of Data Preparation</b>	<ul style="list-style-type: none"> <li>• Apply Structured Query Language to develop profiles from premium and loss data summaries. <ul style="list-style-type: none"> <li>➤ Summarize an Earned Premium Table to a Policy-Term Summary</li> <li>➤ Summarize a Policy-Level Loss File to a Book-Level Summary</li> <li>➤ Join the Policy Table to the Loss Table</li> <li>➤ Join a Table Showing Demographic Information to the Premium and Loss Tables</li> <li>➤ Create Profiles Using the Information From All Tables</li> </ul> </li> </ul>

**Assignment 5: Regulations and Privacy Issues**

<b>MODULE TITLE</b>	<b>LEARNING OBJECTIVES</b>
<b>Data Regulation</b>	<ul style="list-style-type: none"> <li>• Describe the fundamental concepts associated with government data regulation. <ul style="list-style-type: none"> <li>➤ U.S. Data Regulations</li> <li>➤ U.S. State Laws</li> <li>➤ International Legislation</li> </ul> </li> </ul>

**Assignment 6: Data Tools and Exploratory Visualization**

<b>MODULE TITLE</b>	<b>LEARNING OBJECTIVES</b>
<b>Exploratory Data Analysis and Data Transformation</b>	<ul style="list-style-type: none"> <li>• Summarize these aspects of exploratory data analysis (EDA): Uses of EDA; Role of metadata in EDA; Data transformations identified through EDA <ul style="list-style-type: none"> <li>➤ Exploratory Data Analysis</li> <li>➤ Data Transformation</li> </ul> </li> </ul>

<p><b>Characterizing Data with Univariate Displays</b></p>	<ul style="list-style-type: none"> <li>• Describe the use of univariate descriptive statistics and displays, and some basic techniques for meaningful data characterization. <ul style="list-style-type: none"> <li>➤ Basic Descriptive Statistics</li> <li>➤ Graphs, Tables, and Charts</li> <li>➤ Displaying and Assessing Time Series Data</li> <li>➤ Bucketing for Categorical Variables</li> </ul> </li> </ul>
<p><b>Identifying and Treating Data Anomalies</b></p>	<ul style="list-style-type: none"> <li>• Describe these aspects of data anomalies: Types of data anomalies; Methods to detect anomalies; Adjustments to data to reduce the impact on analysis <ul style="list-style-type: none"> <li>➤ Defining Types of Data Anomalies</li> <li>➤ Detecting Data Anomalies</li> <li>➤ Adjustments to Data</li> </ul> </li> </ul>
<p><b>Using Multivariate Summaries and Displays</b></p>	<ul style="list-style-type: none"> <li>• Describe the use of multivariate summaries and displays to analyze data, detect outliers, and/or formulate preliminary hypotheses. <ul style="list-style-type: none"> <li>➤ Pivot Tables</li> <li>➤ Contingency Tables</li> <li>➤ Linear and Nonlinear Correlations</li> <li>➤ Scatterplots and Correlations</li> <li>➤ Heat Maps</li> </ul> </li> </ul>
<p><b>Visualization Methods</b></p>	<ul style="list-style-type: none"> <li>• Describe these aspects of data visualization: Data preparation for visualization; Basic concepts and methods of data visualization <ul style="list-style-type: none"> <li>➤ Data Preparation for Visualization</li> <li>➤ Basic Concepts and Methods of Data Visualization</li> </ul> </li> </ul>
<p><b>Data Visualization Displays</b></p>	<ul style="list-style-type: none"> <li>• Explain how to visualize data using various types of displays. <ul style="list-style-type: none"> <li>➤ Tables</li> <li>➤ Dashboards</li> <li>➤ Charts and Graphs</li> <li>➤ Maps</li> </ul> </li> </ul>

## External Readings – Exploratory Data Analysis

*Exploratory Data Analysis With R - Roger Peng*

TOPICS	LEARNING OBJECTIVES
<b>Chapter 3,4 – Introduction to R and Managing Data Frames</b>	<ul style="list-style-type: none"> <li>• Introduction</li> <li>• Choose functions in the dplyr function to manage data frames and to apply basic data screening procedures</li> </ul>
<b>Chapter 5 – Exploratory Data Analysis Checklist</b>	<ul style="list-style-type: none"> <li>• Construct an exploratory data analysis using the 10-point checklist</li> </ul>
<b>Chapter 6,7 – Graphics and Exploratory Graphs</b>	<ul style="list-style-type: none"> <li>• Apply fundamental principles of analytic graphics</li> <li>• Develop simple summaries and exploratory graphs that optimize data visualization at the beginning stages of data analysis</li> </ul>
<b>Chapter 8-16 - Plotting Systems</b>	<ul style="list-style-type: none"> <li>• Compare the base, lattice, and ggplot2 plotting systems</li> <li>• Create plots on graphic devices including screens and files</li> <li>• Create graphics on the screen device in the base plotting system, including annotation, regression lines and multiple plots</li> <li>• Design plots in R that optimize the use of color and transparency, including applications to large datasets</li> <li>• Apply the qplot function in ggplot2 for elegant expression of plot components</li> <li>• Apply the key components of a ggplot2 plot to build plots in layers, including data frames, aesthetic mappings, geoms, facets, stats, scales, and coordinate system</li> </ul>
<b>Chapter 17 - Data Analysis Case Study: Changes in Fine Particle Air Pollution in the US</b>	<ul style="list-style-type: none"> <li>• Conduct an exploratory data analysis in R using raw data</li> </ul>

*Visualizing Data - William Cleveland*

TOPICS	LEARNING OBJECTIVES
<b>Chapter 2 - Univariate Data</b>	<ul style="list-style-type: none"> <li>Summarize the characteristics and purposes of visualizing univariate data</li> <li>Construct Q-Q plots and normal plots to evaluate data distributions</li> <li>Develop fits and residuals, spread of fits and residuals, quantile plots of residuals and spread location plot</li> <li>Evaluate when to use log scale for displaying data</li> <li>Apply transformations of non-normal variables to approximate normality</li> <li>Apply robust estimation to mitigate distortion from outliers in the data</li> </ul>

*Show Me the Numbers - Stephen Few*

<b>Chapter 1,2 – Introduction</b>	<ul style="list-style-type: none"> <li>Explain how well-designed quantitative tables and graphs communicate important information effectively</li> <li>Compare the types of data and relationships included in quantitative stories</li> <li>Distinguish these summarization measures: average, variation, correlation, ratio</li> </ul>
<b>Chapter 3,4,8 – Tables vs. Graphs, Table Design</b>	<ul style="list-style-type: none"> <li>Evaluate the features of tables and graphs for their optimal use scenarios</li> <li>Classify unidirectional and bidirectional table structures that illustrate quantitative-to-categorical and quantitative-to-quantitative relationships</li> <li>Show how the thoughtful design of information and support components lead to clear and efficient communication with tables</li> </ul>
<b>Chapter 5- Visual Perception and Graphical Communication</b>	<ul style="list-style-type: none"> <li>Explain how the eye and brain process signals in the mechanics of sight</li> <li>Illustrate how the attributes of form, color, and spatial position affect preattentive processing</li> <li>Apply preattentive attributes to design tables and graphs with optimal visual emphasis</li> <li>Summarize the Gestalt principles of visual perception and how we group objects in particular ways</li> </ul>
<b>Chapter 6- Fundamental Variations of Graphs</b>	<ul style="list-style-type: none"> <li>Examine how the fundamental variations of graphs and corresponding quantitative relationships can be paired with optimal visual components and techniques</li> </ul>



<p><b>Chapter 7, 9, 10, 11 – Graph Design</b></p>	<ul style="list-style-type: none"> <li>• Explain how communication-oriented design supports the objectives of highlighting and organizing information</li> <li>• Apply this fundamental principle of quantitative communication to graph design: visual correspondence to quantity</li> <li>• Compare options in primary data component design for graph construction for these components: points, bars, lines, boxes</li> <li>• Design these secondary data components for optimal communication in graphs: trend lines, reference lines, annotations, scales, tick marks, grid lines, legends</li> <li>• Recommend graph design strategies to address presentation of multiple variables including: multiple units of measure in a single graph, combining multiple graphs in a series</li> </ul>
<p><b>Chapter 12 – Silly Graphs that are Best Forsaken</b></p>	<ul style="list-style-type: none"> <li>• Explain why these types of graphs don't communicate effectively: donut charts, radar charts, area charts, circle charts, unit charts, funnel charts, waterfall charts</li> </ul>
<p><b>Chapter 13 - Telling Compelling Stories with Numbers</b></p>	<ul style="list-style-type: none"> <li>• Apply these principles and practices in order to create a compelling statistical narrative: simple, seamless, informative, true, contextual, familiar, concrete, personal, emotional, actionable, sequential</li> </ul>
<p><b>Chapter 14 - The Interplay of Standards and Numbers</b></p>	<ul style="list-style-type: none"> <li>• Explain how the embrace of standards for the visual display of quantitative information eliminates bad choices and creates time for innovative options</li> </ul>

**"Infovis and Statistical Graphics: Different Goals, Different Looks", Andrew Gellman and Antony Unwin**

<p><b>Infovis and Statistical Graphics</b></p>	<ul style="list-style-type: none"> <li>• Explain how the different goals of Statistical visualization and Information graphics affect design of graphs</li> <li>• Distinguish examples of flaws and strengths from Visualization projects that received praise</li> <li>• Describe an example of a graph that displays benefits of both statistical graphs and Infoviz</li> </ul>
--	---

## Assignment 7: Insurance Applications

MODULE TITLE	LEARNING OBJECTIVES
<b>Fundamentals of Modeling Data for Insurance Applications</b>	<ul style="list-style-type: none"> <li>• Describe the fundamental applications of data mining statistical tools for data preparation.               <ul style="list-style-type: none"> <li>➤ Overview of Potential Modeling Methods</li> <li>➤ Screening of Variables for Inclusion in the Final Modeling Dataset</li> <li>➤ Transformation of Variables, Including Binning</li> <li>➤ Identification and Treatment of Missing Values</li> <li>➤ Identification of Outliers and Errors</li> <li>➤ Separation of the Data Into Training and Holdout Data</li> <li>➤ Testing of Models</li> </ul> </li> </ul>
<b>Creating Datasets for Underwriting Models</b>	<ul style="list-style-type: none"> <li>• Describe the fundamental concepts and challenges associated with preparing modeling data for underwriting applications.               <ul style="list-style-type: none"> <li>➤ Identify Target Variable</li> <li>➤ Make Insurance Adjustments</li> <li>➤ Identify Data Sources</li> <li>➤ Gather Metadata</li> <li>➤ Join Data From Different Sources</li> <li>➤ Perform Exploratory Data Analysis and Data Cleansing</li> <li>➤ Transform Variables</li> <li>➤ Identify Potential Independent Variables</li> <li>➤ Perform Data Reduction</li> <li>➤ Create Training/Test/Holdout Samples</li> <li>➤ Document Data Preparation Work</li> </ul> </li> </ul>
<b>Creating Datasets for Claims Models</b>	<ul style="list-style-type: none"> <li>• Describe the fundamental concepts and challenges associated with preparing modeling data for claims applications.               <ul style="list-style-type: none"> <li>➤ Data for Claims Triage Modeling</li> <li>➤ Data for Claims Fraud Modeling</li> <li>➤ Data for Next-Best-Action Modeling</li> </ul> </li> </ul>