

Learning Objectives for Data Concept and Visualization

Assignment 1: Data Quality

MODULE TITLE	LEARNING OBJECTIVES
Concept and Impact of Data Quality	<ul style="list-style-type: none"> Summarize concepts of data quality. Understand and describe the impact of data on actuarial work and projects.
Data Quality Principles	<ul style="list-style-type: none"> 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.
Data Governance	<ul style="list-style-type: none"> Concepts, roles and responsibilities, committees, tools
Data Documentation: Metadata Terminology	<ul style="list-style-type: none"> Describe these aspects of data documentation: <ul style="list-style-type: none"> ➤ Data and metadata terminology ➤ Relationship between data documentation and data governance ➤ Types and uses of metadata for data scientists
Aggregate Insurance and Statistical Data	<ul style="list-style-type: none"> Explain the regulator and business needs for statistical data
Range of Weight: 0-10 %	

Assignment 2: Sources of Data

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

	<ul style="list-style-type: none"> • Understand timing of collection and updating. Understand how quality can change over time.
The Value of Statistical Plan Data	<ul style="list-style-type: none"> • Understand why insurance companies produce statistical files, typical attributes in stat files and the advantages and disadvantages of using stat files over operational data.
Statistical Plans	<ul style="list-style-type: none"> • Describe the two basic types of statistical plans and their use in insurers' predictive modeling and ratemaking
Insurance Date Fields and Amount Fields in Statistical Plans	<ul style="list-style-type: none"> • Explain the role of the statistical agent and how the date and amount fields are used in the statistical plans.
Classification or Rating Variable Fields and Exposure Data Elements in Statistical Plans	<ul style="list-style-type: none"> • Describe the following statistical plan elements by line of business. <ul style="list-style-type: none"> ➤ Classification and rating variable fields ➤ Exposure
External Sources of Noninsurance Data	<ul style="list-style-type: none"> • Understand how to access and the uses of external sources of data including: <ul style="list-style-type: none"> ➤ Demographic information ➤ Customer Financial Information ➤ Business Financial Information ➤ Behavioral data ➤ Driving Records and Motor Vehicle Reports ➤ Government sources • 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. • Understand various derived attributes. Understand if there is a clear way to merge the data into databases used for analysis.
External Sources of Data – Insurance Specific	<ul style="list-style-type: none"> • Understand External sources of Insurance Specific data such as: <ul style="list-style-type: none"> ➤ Historical claims reports ➤ Industry trend factors ➤ Loss Development factors ➤ Data available in NCCI Statistical Bulletin
External Data Sources: United States Census Data	<ul style="list-style-type: none"> • Understand the two basic formats into which data from the United States Census bureau is organized.
Range of Weight: 10-20 %	

Assignment 3: Accessing Raw Data

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

Data Profiling	<ul style="list-style-type: none"> • Demonstrate ability to profile data including: inspecting rows of data, read data catalogs and metadata. Create descriptive statistics and graphs that profile the data.
Messy Data	<ul style="list-style-type: none"> • Detect and remediate missing, miscoded and anomalous data. Understand sampling bias and clustering of values.
Testing Data	<ul style="list-style-type: none"> • Work with small test data and create sample data using simple filters. Be able to sample from related tables.
Range of Weight: 10-20 %	

Assignment 4: Working with Data

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

Working With Regular Expressions in String Processing	<ul style="list-style-type: none"> • Understand the use of common regular expressions for pattern matching. <ul style="list-style-type: none"> ➤ Basic Regular Expressions ➤ Application of Regular Expressions
Using Hash Tag Functions	<ul style="list-style-type: none"> • Explain how to use hash functions with databases. <ul style="list-style-type: none"> ➤ Common Hash Functions ➤ Hash Functions for Equality Testing ➤ Hash Functions for Hash Tables ➤ Hash Functions for Data Segmentation
Insurance Applications of Data Preparation	<ul style="list-style-type: none"> • Apply Structured Query Language to develop profiles from premium and loss data summaries. <ul style="list-style-type: none"> ➤ Summarize an Earned Premium Table to a Policy-Term Summary ➤ Summarize a Policy-Level Loss File to a Book-Level Summary ➤ Join the Policy Table to the Loss Table ➤ Join a Table Showing Demographic Information to the Premium and Loss Tables ➤ Create Profiles Using the Information From All Tables
Range of weight: 0-10 %	

Assignment 5: Regulations and Privacy Issues

MODULE TITLE	LEARNING OBJECTIVES
Data Regulation	<ul style="list-style-type: none"> • Describe the fundamental concepts associated with government data regulation. <ul style="list-style-type: none"> ➤ U.S. Data Regulations ➤ U.S. State Laws ➤ International Legislation
Range of weight: 0-10 %	

Assignment 6: Data Tools and Exploratory Visualization

MODULE TITLE	LEARNING OBJECTIVES
Exploratory Data Analysis and Data Transformation	<ul style="list-style-type: none"> • 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"> ➤ Exploratory Data Analysis ➤ Data Transformation
Characterizing Data with Univariate Displays	<ul style="list-style-type: none"> • Describe the use of univariate descriptive statistics and displays, and some basic techniques for meaningful data characterization. <ul style="list-style-type: none"> ➤ Basic Descriptive Statistics ➤ Graphs, Tables, and Charts ➤ Displaying and Assessing Time Series Data ➤ Bucketing for Categorical Variables
Identifying and Treating Data Anomalies	<ul style="list-style-type: none"> • 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"> ➤ Defining Types of Data Anomalies ➤ Detecting Data Anomalies ➤ Adjustments to Data
Using Multivariate Summaries and Displays	<ul style="list-style-type: none"> • Describe the use of multivariate summaries and displays to analyze data, detect outliers, and/or formulate preliminary hypotheses. <ul style="list-style-type: none"> ➤ Pivot Tables ➤ Contingency Tables ➤ Linear and Nonlinear Correlations ➤ Scatterplots and Correlations ➤ Heat Maps
Visualization Methods	<ul style="list-style-type: none"> • Describe these aspects of data visualization: Data preparation for visualization; Basic concepts and methods of data visualization <ul style="list-style-type: none"> ➤ Data Preparation for Visualization ➤ Basic Concepts and Methods of Data Visualization
Data Visualization Displays	<ul style="list-style-type: none"> • Explain how to visualize data using various types of displays. <ul style="list-style-type: none"> ➤ Tables ➤ Dashboards ➤ Charts and Graphs ➤ Maps
Range of weight: 0-10 %	

External Readings – Exploratory Data Analysis

Assignment 7: Exploratory Data Analysis With R - Roger Peng

TOPICS	LEARNING OBJECTIVES
Chapter 3,4 – Introduction to R and Managing Data Frames	<ul style="list-style-type: none"> • Introduction • Choose functions in the dplyr function to manage data frames and to apply basic data screening procedures
Chapter 5 – Exploratory Data Analysis Checklist	<ul style="list-style-type: none"> • Construct an exploratory data analysis using the 10-point checklist
Chapter 6,7 – Graphics and Exploratory Graphs	<ul style="list-style-type: none"> • Apply fundamental principles of analytic graphics • Develop simple summaries and exploratory graphs that optimize data visualization at the beginning stages of data analysis
Chapter 8-16 - Plotting Systems	<ul style="list-style-type: none"> • Compare the base, lattice, and ggplot2 plotting systems • Create plots on graphic devices including screens and files • Create graphics on the screen device in the base plotting system, including annotation, regression lines and multiple plots • Design plots in R that optimize the use of color and transparency, including applications to large datasets • Apply the qplot function in ggplot2 for elegant expression of plot components • 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
Chapter 17 - Data Analysis Case Study: Changes in Fine Particle Air Pollution in the US	<ul style="list-style-type: none"> • Conduct an exploratory data analysis in R using raw data
Range of weight: 10-20 %	

Assignment 8: Visualizing Data - William Cleveland

TOPICS	LEARNING OBJECTIVES
Chapter 2 - Univariate Data	<ul style="list-style-type: none"> • Summarize the characteristics and purposes of visualizing univariate data • Construct Q-Q plots and normal plots to evaluate data distributions • Develop fits and residuals, spread of fits and residuals,

	<p>quantile plots of residuals and spread location plot</p> <ul style="list-style-type: none"> • Evaluate when to use log scale for displaying data • Apply transformations of non-normal variables to approximate normality • Apply robust estimation to mitigate distortion from outliers in the data
Range of weight: 0-10 %	

Assignment 9: *Show Me the Numbers* - Stephen Few

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

Chapter 7, 9, 10, 11 – Graph Design	<ul style="list-style-type: none"> • Explain how communication-oriented design supports the objectives of highlighting and organizing information • Apply this fundamental principle of quantitative communication to graph design: visual correspondence to quantity • Compare options in primary data component design for graph construction for these components: points, bars, lines, boxes • Design these secondary data components for optimal communication in graphs: trend lines, reference lines, annotations, scales, tick marks, grid lines, legends • 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
Chapter 12 – Silly Graphs that are Best Forsaken	<ul style="list-style-type: none"> • Explain why these types of graphs don't communicate effectively: donut charts, radar charts, area charts, circle charts, unit charts, funnel charts, waterfall charts
Chapter 13 - Telling Compelling Stories with Numbers	<ul style="list-style-type: none"> • Apply these principles and practices in order to create a compelling statistical narrative: simple, seamless, informative, true, contextual, familiar, concrete, personal, emotional, actionable, sequential
Chapter 14 - The Interplay of Standards and Numbers	<ul style="list-style-type: none"> • Explain how the embrace of standards for the visual display of quantitative information eliminates bad choices and creates time for innovative options
Range of weight: 10-20 %	

Assignment 10: "Infovis and Statistical Graphics: Different Goals, Different Looks", Andrew Gellman and Antony Unwin

Infovis and Statistical Graphics	<ul style="list-style-type: none"> • Explain how the different goals of Statistical visualization and Information graphics affect design of graphs • Distinguish examples of flaws and strengths from Visualization projects that received praise • Describe an example of a graph that displays benefits of both statistical graphs and Infoviz
Range of weight: 0-10 %	

Assignment 11: Insurance Applications

MODULE TITLE	LEARNING OBJECTIVES
Fundamentals of Modeling Data for Insurance Applications	<ul style="list-style-type: none"> • Describe the fundamental applications of data mining statistical tools for data preparation. <ul style="list-style-type: none"> ➤ Overview of Potential Modeling Methods ➤ Screening of Variables for Inclusion in the Final Modeling Dataset ➤ Transformation of Variables, Including Binning ➤ Identification and Treatment of Missing Values ➤ Identification of Outliers and Errors ➤ Separation of the Data Into Training and Holdout Data ➤ Testing of Models
Creating Datasets for Underwriting Models	<ul style="list-style-type: none"> • Describe the fundamental concepts and challenges associated with preparing modeling data for underwriting applications. <ul style="list-style-type: none"> ➤ Identify Target Variable ➤ Make Insurance Adjustments ➤ Identify Data Sources ➤ Gather Metadata ➤ Join Data From Different Sources ➤ Perform Exploratory Data Analysis and Data Cleansing ➤ Transform Variables ➤ Identify Potential Independent Variables ➤ Perform Data Reduction ➤ Create Training/Test/Holdout Samples ➤ Document Data Preparation Work
Creating Datasets for Claims Models	<ul style="list-style-type: none"> • Describe the fundamental concepts and challenges associated with preparing modeling data for claims applications. <ul style="list-style-type: none"> ➤ Data for Claims Triage Modeling ➤ Data for Claims Fraud Modeling ➤ Data for Next-Best-Action Modeling
Range of weight: 0-10 %	