

DSO 545: HW 1 **DUE February 1**

Learning objectives and outcomes:

1. Conduct Exploratory Business Data Analysis R
2. The assignment offers comprehensive coverage of R's base installation.
3. The assignment weaves the technical content into realistic business scenarios and focuses on using R as a decision-making tool.
4. The assignment cases help you progress from a basic understanding to mastery of each application, empowering you to perform an analysis of each case with confidence in R.
5. With more information from data analysis, you will be equipped to make superior decisions and outperform the competition.
6. Your mastery of essential skills of creating and communicating data analysis for improved decision making will enhance your career and make numbers fun.

Instructions:

Submit(one per team):

1. **Word file with answers inserted below each question**
2. **A script file(R code file) that contains properly labeled/numbered chunks of R code for each question**

Case 1



Background and Objective: This case is based on the Air Travel Consumer Reports. The Air Travel Consumer Report is a monthly product of the Department of Transportation's Office of Aviation Enforcement and Proceedings (OAEP). Anyone who travels by air knows that occasional problems are inevitable. Flights can be delayed or cancelled due to weather conditions, mechanical problems, or labor strikes, and baggage can be lost, delayed, damaged, or pilfered. Given that many airlines are now charging for bags, issues with baggage are particularly annoying. You're standing at a baggage carousel for what seems like forever when the steady flow of bags onto the conveyor belt slows to a trickle, then stops. Your bags are nowhere in sight. Or your bags do show up, but look like an angry gorilla has been throwing them around his cage for sport. Baggage problems can have a serious impact on customer loyalty, and can be costly to the airlines (airlines often have to deliver bags).

Air carriers report flight delays, cancellations, overbookings, late arrivals, baggage complaints, and other operating statistics to the U.S. government, which compiles the data and reports it to the public. This data is available from the U.S. Department of Transportation, "Air Travel Consumer Report," the Office of Aviation Enforcement and Proceedings, Aviation Consumer Protection Division (see <http://www.transportation.gov/airconsumer/July-2015-air-travel-consumer-report> for most recent reports). The data for baggage complaints and enplaned passengers cover domestic travel only.

Do some airlines do a better job of handling baggage? Compare the baggage complaints for three airlines: American Eagle, Hawaiian, and United. Which airline has the best record? The worst? Are complaints getting better or worse over time? Are there other factors, such as destinations, seasonal effects or the volume of travelers that affect baggage performance?

Data: The data set in **Baggage.csv** contains monthly observations from 2004 to 2010 for United Airlines, American Eagle, and Hawaiian Airlines. The variables in the data set include:

Baggage = the total number of passenger complaints for theft of baggage contents, or for lost, damaged, or misrouted luggage for the airline that month

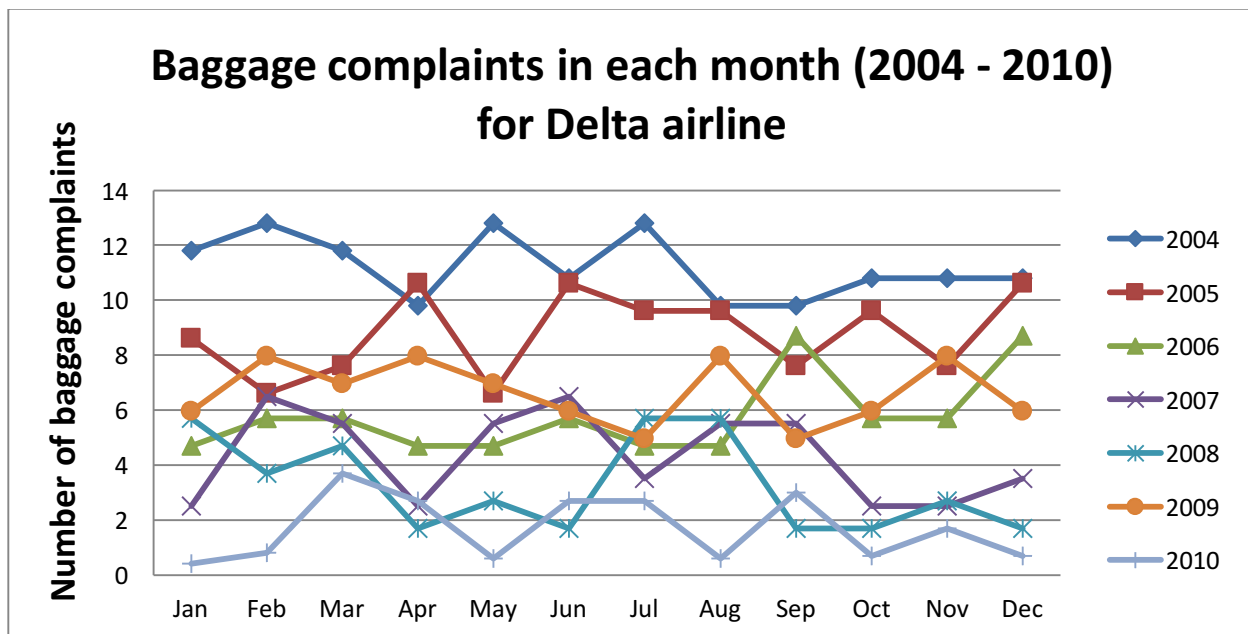
Scheduled = the total number of flights scheduled by that airline that month

Cancelled = the total number of flights cancelled by that airline that month

Enplaned = the total number of passengers who boarded a plane with the airline that month

Questions:

1. Explore baggage complaints over time: create 3 time series plots for the variable **Baggage** by Date for each of the airlines separately.
2. Briefly describe what patterns, if any, you see in the plots in 1.
3. To compare each month's data in different years, superimpose each year's **Baggage** by month on a single graph for each airline. Your answer consists of 3 time series plots. The plot below is a hypothetical example of what each airline's plot should look like:



4. Briefly describe what patterns, if any, you see in the plots in 3.
5. To better compare the baggage complaints for three airlines, plot all three airline Baggage data by Date on one graph.
6. Based on the graph in question 5., do some airlines have better baggage handling practices?
7. Based on the graph in question 5., which airline has the best record? The worst?
8. Based on the graph in question 5., are complaints getting better or worse over time?
9. Are the conclusions, you have drawn based on the graphs of the raw data you created, accurate? Are there any potential factors that may distort your conclusions and should be taken into consideration?

10. Report the average of scheduled flights and the average of enplaned passengers by airline.
11. What insights, ideas, and concerns does the data in the table in 10. provide you with?
12. Create a KPI that adjusts the total number of passenger complaints for size in the following way: **Baggage % = Baggage / Enplaned × 100 %**. Display average **Baggage %** for each airline.
13. Do the results in question 12 support your previous conclusions? Briefly explain.
14. Superimpose all three time series on one graph to display **Baggage %** by Date.
15. In addition to the graph in question 14., would plotting each series on a separate graph be beneficial and why? Create a graph to support your answer.
16. Based on the analysis of KPI **Baggage %**, have any of your conclusions drawn in questions 6. - 8. changed? Briefly discuss.
17. Superimpose time series plots of monthly averages of **Baggage %** by time for the three airlines.
18. Discuss common patterns all three time series exhibit in question 17.
19. Create a timeplot of **Baggage %**, add average line of **Baggage %** and a Trendline (regression line) of monthly average **Baggage %**'s for each airline. Hint: total of 3 charts; each chart 3 superimposed trajectories.
20. Prepare a brief (one paragraph) executive summary of your findings.

Case 2

CEO compensation has received great attention in the recent economic recession. File **CEOcompensation.xlsx** contains data on the most highly compensated CEOs in 2008, according to a Forbes Web site (the following link helps you find most recent data; for this assignment data is posted on blackboard!):

https://www.forbes.com/lists/2012/12/ceo-compensation-12_land.html

The data set on blackboard includes some personal data about each CEO: components (columns K – N) of the CEO's total compensation for 2008, the CEO's total 5-year compensation, and the value of company shares owned by the CEO. For those CEOs with tenure at least six years, it also shows some six-year values, including a performance versus pay ranking (1 is best) in Column U. Finally the last 2 columns indicate the CEO's total return versus tenure and how this compares to the market.

A Primer on CEO Compensation

If a normal person gets a salary of \$80,000 in a given year, that's about the end of the story. However, CEO compensation is considerably more complex. Each CEO receives a base salary (column K) and an incentive bonus (column L), the latter decided by negotiation. A CEO can also receive "other compensations" (column M), including vested restricted stock grants, LTIP (long-term incentive plan) payouts, and perks. However, the big difference between CEOs and the rest of us is the granting of stock options. A stock option allows a CEO to purchase company stock at a fixed stated price during a certain period of time, often 10 years. If the price of the

company stock increases during that period, the CEO can then exercise the stock options by buying the stock at the low fixed price and selling it at the high current price, thereby making a windfall. This explains the huge stock gains in column N for several of the CEOs in the data table. They evidently exercised at least some of their options in 2008. In a sense, these huge stock gains for some CEOs overstate their compensation for 2008. They had been holding these valuable stock options for years, but their gains showed up only in 2008 when they exercised the options.

Data: CEOcompensation.txt and IndustryMedians.csv

Note that all monetary values in the data file are in \$ millions. Descriptions of the variables are indicated below:

Variable	Description
2008 Salary	Base salary
2008 Bonus	A contractual incentive for high company performance
2008 Other	Includes vested restricted stock grants, LTIP (long-term incentive plan) payouts, and perks
2008 Stock gains	Windfall from exercising stock options in 2008
5-year compensation total	Or for as many years as CEO if this is less than 5
Shares owned (\$ millions)	Value of company stock owned
6-year average compensation	Columns Q-U are reported by Forbes to gauge the performance of the company over a 6-year period to the CEO's compensation over this period, relative to the market. These are reported only for CEOs with long enough tenure in their present positions.
6-year annual total return	Annualized stock return (including dividends) over the last 6 years
6-year return relative to industry	Performance of the company's stock relative to the industry (a score of 100 = the industry average).
6-year return relative to market	Performance of the company's stock relative to the S&P 500 (a score of 100 = the S&P 500).
Performance vs pay rank	1 best, 179 worst, although it is not absolutely clear how this ranking was made. Ranks are given only to CEOs who have a six-year tenure and six-year compensation history.
Total return during tenure	Annualized stock return (including dividends) for the tenure of the CEO of from the time of the IPO or available stock history.
Relative to market	The ending value of \$100 invested in the stock, divided by the ending value of \$1 invested in the S&P 500 (a score of 100 = the S&P 500).

Questions:

1. What is the number of female CEOs? (Answer is an integer)
2. What is the age of a youngest CEO? (Answer is an integer)
3. What is the age of the oldest CEO? (Answer is an integer)
4. What is the average age of a CEO? (Use two decimal digit precision)
5. What is the total CEO 2008 salary? (Use two decimal digit precision)

6. How many CEOs have joined a company as a CEO? (Hint: CEOs can always be founders. Founders can't always be CEOs)
7. What is the average amount of time a CEO worked for a company before becoming a CEO? (Use two decimal digit precision)
8. Which industry in the data set has largest number CEO's?
9. What is the average CEO 2008 Compensation? Note that 2008 compensation for a CEO consists of a total four components: Salary, Bonus, other (including vested restricted stock grants, LTIP (long-term incentive plan) payouts, and perks), and stock gains. (Use two decimal digit precision)
10. Which CEO did get paid the largest compensation amount in 2008?
11. What is the corresponding amount? (Use two decimal digit precision)
12. Which industry does correspond to the second largest total CEO compensation in 2008? (Hint:check sort(), order () functions).
13. Consider the following age groups: [45 – 50), [50 – 55), [55 – 60), [60 – 70), and [70 or more). Analyze age groups by industry and determine which age group corresponds to largest CEO average salary in 2008? Hint: 1. left end point is included; 2. nested **if** helps assign age category

Additional info: File **IndustryMedians.xls** includes median compensation/salary/other/bonus/stock gain values for the various industries, so that you can compare any CEO to these median values in his/her industry. Compare each CEO's total compensation to corresponding median compensations in his/her respective industry by calculating % difference:

$$\% \text{difference} = (\text{CEO's compensation in 2008} - \text{Corresponding industry median compensation in 2008}) / \text{Corresponding industry median compensation in 2008} \times 100\%$$

14. How many CEO's have received 100% or larger compensation relative to their respective median compensation?
15. Is the following formula always true?

$$\text{Total median compensation} = \text{Median Salary} + \text{Median Bonus} + \text{Median Other} + \text{Median Stock Gains}$$