# DSCI 210 – Activity 7.4: Reshaping Tables in Tableau Prep

## Exercise 7.4.1 – Recoding Survey Data using the Stack-Transform-Unstack Trick

The file **health\_survey.csv** contains the responses to a series of health-related questions. Dr. Bergen, Director of the Statistical Consulting Center at WSU, needs you to prepare the attached data for analysis. Please perform the following steps to prepare the required csv file.

**Outline of the Process.**

1. Our first goal is to transform the survey responses to a numeric scale, as shown below. Instead of performing the same transformation on each column, we will use the *Stack + Transform* trick from the previous lecture.

|  |  |
| --- | --- |
| Old Label | New Coded Value |
| “Strongly Disagree” | 1 |
| “Somewhat Disagree” | 2 |
| “Neither Agree nor Disagree” | 3 |
| “Somewhat Agree” | 4 |
| “Strongly Agree” | 5 |

1. Notice that there are several columns that start with the same letters (“F1”, “F1.1”, “F1.2”, etc.). We need to make a new column named *Main Type* that contains the first two letters of the column (e.g. “F1.1” 🡪 “F1”)
2. Once we have converted the responses to a numeric scale and the labels to the the main question types, we will aggregate the coded values by summing the values for each of the *Main Type*s.
3. Finally, you should unstack/split these new aggregated values into one column per *Main Type*.

**Task.** Perform the data management process in Tableau Prep, using the following sequence of screenshots as a guide.

**Start**

![A screenshot of a computer

Description automatically generated]()

**Stack**

A screenshot of a computer

Description automatically generated

**Recode\*2**

A screenshot of a computer

Description automatically generated

**Aggregate**

A screenshot of a computer

Description automatically generated

**Split**

A table of numbers on a white surface

Description automatically generated

**Deliverables.** Save the resulting file as survey\_summary.csv. Submit both the Tableau Prep file and CSV to D2L.

## Exercise 7.4.2 – Visualizing a difference.

In a previous activity, you used JMP to prepare the data and create the following visualization, which illustrates the effect of the designated hitter on the best pitchers. In this exercise, you will create a similar visualization using Tableau Prep and Desktop.

A picture containing chart

Description automatically generated

Figure . Visualization from a previous activity comparing the league-wide min(ERA) across years.

Again, we will be starting with the Teams.csv file. This file contains, for each season, team-by-team statistics. We will focus on the ERA, which measures the average number of runs allowed by each team’s pitchers over a 9-inning game, with a smaller number indicating better pitching + defense.

**Task 1 – Data Management**

Because we are using an overlay for both leagues, *as well as the difference between leagues*, we will need to use reshaping techniques to make this graph. To create the graph, perform the following steps.

1. Load the Teams.csv file in Tableau Prep.
2. *Select* the columns of interest.

**A table with numbers and letters

Description automatically generated**

1. *Filter* the data to only the years after World War II (1946+).

**A table with numbers and letters

Description automatically generated**

1. In this activity, we will be comparing the yearly league-wide ERA, which is computed using

As is often the case, to compute a rate, we need to first use aggregation to compute the numerator and denominator, then divide the results. *Group and aggregate* the data to compute the yearly Total ERA and Total IPouts for each league. You may want to add an extra clean step to inspect the resulting table.

**A table with numbers and letters

Description automatically generated**

1. Now compute the total games pitched and yearly league-wide ERA. Drop the intermediate columns. HINT. A game has 27 outs.

**A screenshot of a table

Description automatically generated A table with numbers and letters

Description automatically generated**

1. Now we want to compute the difference between the yearly *League-wide ERA* for the leagues but currently these values are on different rows. To do this we will need to split the leagues into separate columns, then compute the difference.

**A table with numbers and letters

Description automatically generated A screenshot of a computer

Description automatically generated A screenshot of a table with numbers

Description automatically generated**

1. We are going to save two versions of this file, so you will want to split your flow into two parts, with an output step in each.

**A screen shot of a computer

Description automatically generated**

* 1. The file **League\_ERA\_unstacked** should contain the ERAs (each league + difference) in separate columns.
  2. The file **League\_ERA\_stacked** should have all the ERAs stacked into a single column.

Make sure you run each output step and save your flow file.

**Task 2 – Make two visualizations.**

Next you should open each of the data files from the last step in Tableau Prep, then create the visualization as shown below.

1. Visualization using the stacked file.

A graph of a graph

Description automatically generated with medium confidence

1. Visualization using the unstacked file.

A screenshot of a graph

Description automatically generated

**Deliverables.** Upload all your files (Flow + Desktop + CSV) to D2L.