

# Goodyear study for treating water from reverse osmosis

---

*Douglas Nychka* May 30, 2020

These data are a formatted and modified version of the data collected for the Goodyear Wetlands Pilot Study. It is based on an excel spreadsheet supplied by Kathryn Newhart for the MoWater summer program in data science and waste water treatment.

The folder clean has the following files with the extensions: `.csv` (comma separated text files), `.pdf` (document), `.xlsx` (MS excel spreadsheet), `.R` (script with R code), `.md` (markdown format) and `.rda` (R binary file of formatted data).

## Quick start

Briefly, to load this dataset into an R session and list these data objects (data frames):

```
load("goodyearMoWater.rda")
> ls()
[1] "Bin1"      "Bin2"      "Bin3"      "Bin4"      "Bin5"
[6] "Bin6"      "Bin7"      "Brine"     "goodyear"  "rawField"
[11] "rawFlow"
```

with `goodyear` being the primary data frame for analysis.

To recreate this R version from the source spread sheets

```
source("createRObjects.R")
```

Note that this script will use the intermediate `.csv` files that were created "by hand" from editing and saving the separate spreadsheets in the original data excel file. Also this script uses the function in **createBinData.R** and several other packages. The strategy is to first form separate data frames for each original spreadsheet and then combine them together into a single data frame for easier analysis.

## NOTES

- Many of the variable names have been renamed to be consistent or shortened across the Bin and Brine data sets. See the R script for these changes.
- #N/A missing value code has been replaced by NA in csv files
- Character strings representing the date (time) have been converted to a data object to make it easier to

work with these times. There may be some inconsistencies in these times. For example they may not be in order for the brine and bin samples. E.g. see `diff( Brine$date)` .

- Columns are variables reported in the experiment and rows are times. For the data frame `goodyear` the additional column **ID** is a factor variable indicating the bin or the brine category for the observations.

## Folder contents

The individual files are described below.

FILE	CONTENTS
<b>README.md, README.pdf</b>	This file
<b>BIN1.csv, BIN1.csv, ... BIN7.csv</b>	edited spreadsheets for the 7 bins of the experiment
<b>Brine.csv</b>	edited brine spreadsheet (aka RO)
<b>Goodyear Wetlands Pilot Study Final Report.pdf</b>	Report submitted to the US Department of Interior
<b>GoodyearWetlandsRawDataset.xlsx</b>	Original data as several spreadsheets
<b>createBinData.R</b>	Reads individual Bin csv files and creates a data frame
<b>createROjects.R</b>	Reads csv files, creates individual R data frames for each spreadsheet and compiles a single data frame for the bin data and brine.
<b>goodyearEDA.R</b>	Some exploratory plots of the data set.
<b>goodyearEDA.pdf</b>	Output from applying notebook function in Rstudio to the exploratory script.
<b>goodyearMoWater.rda</b>	R binary file with all the data frames created in <b>createROjects.R</b> .
<b>rawField.csv</b>	Edited csv version of flow data spread sheet
<b>rawFlow.csv</b>	Edited csv version of flow data spread sheet