Andrew O. Finley and Jeffrey W. Doser

Forestry 472: Ecological Monitoring and Data Analysis

To my son,

without whom I should have finished this book two years earlier

Contents

List of Tables

List of Figures

Preface

This text is an introduction to data sciences for Forestry and Environmental students. Understanding and responding to current environmental challenges requires strong quantitative and analytical skills. There is a pressing need for professionals with data science expertise in this data rich era. The McKinsey Global Institute[1] predicts that "by 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions". The Harvard Business Review dubbed *data scientist* "The Sexiest Job of the 21st Century"[2]. This need is not at all confined to the tech sector, as forestry professionals are increasingly asked to assume the role of *data scientists* and *data analysts* given the rapid accumulation and availability of environmental data (see, e.g. **?**). Thomson Nguyen's talk[3] on the difference between a data scientist and a data analyst is very interesting and contains elements relevant to the aim of this text. This aim is to give you the opportunity to acquire the tools needed to become an environmental data analyst. Following **?** a *data analyst* has the ability to make appropriate calculations, convert data to graphical representation, interpret the information presented in graphical or mathematical forms, and make judgements or draw conclusions based on the quantitative analysis of data.

[1] http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

[2] https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

[3] www.import.io/post/data-scientists-vs-data-analysts-why-the-distinction-matters

0

Data

0.1 FEF Tree Biomass Data Set

When thinking about data, we might initially have in mind a modest-sized and uncomplicated data set that serves a fairly specific purpose. For example, in forestry it is convenient to have a mathematical formula that relates a tree's diameter (or some other easily measured attribute) to stem or total biomass (i.e. we cannot directly measure tree biomass without destructive sampling). When coupled with forest inventory data, such formulas provide a means to estimate forest biomass across management units or entire forest landscapes. A data set used to create such formulas includes felled tree biomass by tree component for four hardwood species of the central Appalachians sampled on the Fernow Experimental Forest[4] (FEF), West Virginia **?**. A total of 88 trees were sampled from plots within two different watersheds on the FEF. Hardwood species sampled include *Acer rubrum*, *Betula lenta*, *Liriodendron tulipifera*, and *Prunus serotina*, all of which were measured in the summer of 1991 and 1992. Data include tree height, diameter, as well as green and dry weight of tree stem, top, small branches, large branches, and leaves. Table **??** shows a subset of these data

The size of this dataset is relatively small, there are no missing observations, the variables are easily understood, etc.

0.2 FACE Experiment Data Set

We often encounter data gleaned from highly structured and complex experiments. Such data typically present challenges in organization/storage, exploratory data analysis (EDA), statistical analysis, and interpretation of analysis results. An example data set comes from the Aspen Free-Air Carbon Diox-

[4] http://www.nrs.fs.fed.us/ef/locations/wv/fernow

TABLE 0.1: A subset of the tree biomass data from the FEF.

species	dbh_in	height_ft	stem_green_kg	leaves_green_kg
Acer rubrum	6.0	48.0	92.2	16.1
Acer rubrum	6.9	48.0	102.3	12.9
Acer rubrum	6.4	48.0	124.4	16.5
Acer rubrum	6.5	49.0	91.7	12.0
Acer rubrum	7.2	51.0	186.2	22.4
Acer rubrum	3.1	40.0	20.8	0.9
Acer rubrum	2.0	30.5	5.6	1.0
Acer rubrum	4.1	50.0	54.1	6.1
Acer rubrum	2.4	28.0	10.2	2.5
Acer rubrum	2.7	40.4	20.2	1.6

ide Enrichment[5] (FACE) Experiment conducted from 1997-2009 on the Harshaw Experimental Forest[6] near Rhinelander, Wisconsin. The Aspen FACE Experiment was a multidisciplinary study that assessed the effects of increasing tropospheric ozone and carbon dioxide concentrations on the structure and functioning of northern forest ecosystems. The design provided the ability to assess the effects of these gasses alone (and in combination) on many ecosystem attributes, including growth, leaf development, root characteristics, and soil carbon. The data set considered here comprises annual tree height and diameter measurements from 1997 to 2008 for *Populus tremuloides*, *Acer saccharum*, and *Betula papyrifera* grown within twelve 30 meter diameter rings in which the concentrations of tropospheric ozone and carbon dioxide were controlled **?**. Because there was no confinement, there was no significant change in the natural, ambient environment other than elevating these trace gas concentrations. Although the basic individual tree measurements are similar to those in the FEF data set we saw in Section~**??**, (i.e., height and diameter), the study design specifies various tree species clones, varying gas treatments, and treatment replicates. Further, because these are longitudinal data, (measurements were recorded over time) the data set presents many missing values as a result of tree mortality. Table **??** contains the first five records as well as 5 more randomly selected records in the data set. Here, a row identifies each tree's experimental assignment, genetic description, and growth over time.

Notice that several height measurements in 2008 contain missing data. If all year measurements were shown, we would see much more missing data. Also, notice that this data set is substantially larger than the FEF data set with 912 rows and 39 columns of data in the full data set.

[5] http://www.nrs.fs.fed.us/disturbance/climate_change/face
[6] http://www.nrs.fs.fed.us/ef/locations/wi/rhinelander/