Lab 2 Homework

Kyle Landolt

February 6, 2015

Setting directory, packages, and data.frame

setwd("~/GEOG415/klandolt/homework2")  
library(ggplot2)  
library(dplyr)

##   
## Attaching package: 'dplyr'  
##   
## The following object is masked from 'package:stats':  
##   
## filter  
##   
## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(tidyr)  
library(magrittr)

##   
## Attaching package: 'magrittr'  
##   
## The following object is masked from 'package:tidyr':  
##   
## extract

precip <- read.csv('data/precip.csv')  
rwi <- read.csv('data/rwi.csv')

# Part 1. Analysis of El Malpais data

What do the tree ring data tell us about monthly precipitation?

Write up a short report (~ 2 paragraphs) answering the following questions:

* Describe the monthly rainfall pattern.
* During what months is tree ring width correlated with rainfall?
* How appropriate is (linear) correlation as a description of the relationship between tree ring width and monthly rainfall?
* If you had to create a 12 month period that was correlated with tree ring width, what 12 months would those be?

In answering these questions, I want to see:

* A bar chart or line graph showing average rainfall each month.

temp.data <- gather(data=precip, key=month, value=precip, -year)  
temp.data$precip[temp.data$precip<0]<- NA  
tidy.precip <- arrange(temp.data, year, month)  
temp.data <- mutate(tidy.precip, date.str = paste(year, month, '01'))  
temp.data <- mutate(temp.data, date=as.Date(date.str, '%Y %B %d'))  
temp.data <- group\_by(tidy.precip, month)  
temp.data <- summarize(temp.data, precip=mean(precip, na.rm=TRUE))  
ggplot(temp.data, aes(month, precip)) + geom\_bar(stat='identity') + labs(y='Precipitation (in)', x='Month') + ggtitle('Monthly Precipitation') + theme(plot.title = element\_text(lineheight=.8, face="bold", size=15))

* Two scatterplots of tree ring width and precipitation, one with the current January, and one with the current August. Use these to assess whether the relationship is approximately linear or not at all linear.

rwi.precip <- left\_join(precip, rwi, by='year')  
ggplot(data=rwi.precip) + geom\_point(aes(x=rwi, y=Jan)) + labs(x='Ring Width Index (RWI)', y= 'Month of January') + ggtitle('Scatterplot of Ring Width Index to the month of January') + theme(plot.title = element\_text(lineheight=.8, face="bold", size=15))

## Warning: Removed 10 rows containing missing values (geom\_point).

rwi.precip <- left\_join(precip, rwi, by='year')  
ggplot(data=rwi.precip) + geom\_point(aes(x=rwi, y=Aug)) + labs(x='Ring Width Index (RWI)', y= 'Month of August') + ggtitle('Scatterplot of Ring Width Index to the month of August') + theme(plot.title = element\_text(lineheight=.8, face="bold", size=15))

## Warning: Removed 10 rows containing missing values (geom\_point).

* Two (or one) plots showing the correlation at every month between tree ring width and precipitation, for both the current year and the previous year. If you can get both current and previous (lagged) year on one plot, then great.

tidy.precip <- gather(data=precip, key=month, value=precip, -year) %>%  
 arrange(year,month) %>%  
 filter(precip >= 0)  
rwi.precip <- left\_join(rwi, tidy.precip, by='year') %>% filter(year >= 1895)  
cor <- rwi.precip %>% group\_by(month) %>% summarise(r=cor(rwi,precip))  
cor1 <- rwi.precip %>% group\_by(month) %>% summarise(r=cor(rwi, precip, use='complete.obs'))  
corlag <- rwi.precip %>% group\_by(month) %>% summarize(r = cor(rwi, lag(precip),use='complete.obs'))  
totcor <- left\_join(cor, corlag, by='month')  
colnames(totcor) <- c("month", "cor.r", "corlag.r")  
ggplot(data=totcor) + geom\_point(aes(x=month, y=cor.r, color='current year')) + geom\_point(aes(x=month, y=corlag.r, color='lagged year')) + labs(color='Correlation of:') + geom\_abline(intercept=0, slope=0) + labs(x='Month', y= 'Correlation value (r)') + ggtitle('Correlation value \n of lagged year and current year to month') + theme(plot.title = element\_text(lineheight=.8, face="bold", size=15))

Make sure that your charts are properly labeled and are referred to in your answer.

# Part 2: Getting the interpretation of correlation correct.

1. In regard to the first part, which of the following is more correct and explain.

* A positive correlation between tree ring width and January precipitation might happen because January is often (but not always) wetter than the average month.
* A positive correlation between tree ring width and January precipitation might happen if dryer-than-average Januarys are associate with smaller-than-average tree ring widths.
* Would the correlation between the age of a second-hand car and its price be positive or negative? Why? (Antiques are not included.)
* How would including antique cars change your response?

1. True or False, and explain: if the correlation coefficient is 0.90, then 90% of the points are highly correlated.
2. An economist is studying the relationship between household income and teenage pregnancy in the United States. She has data for each of the 50 states, showing the average household income of families with teenagers, and the rates of pregnancy among teenage girls in that state. The correlation is 0.4. Does this give a fair estimate of th e strength of the association between household income and teenage pregnancy? Explain.
3. Three data sets are collected and the correlation coefficient is computed in each case. The variables in the datasets are:

* grade point average in freshman year and sophomore year in college.
* grade point average in freshman year and senior year in college.
* length and weight of two-by-four boards of lumber.

Possible values for correlation coefficients are: -0.5, 0.0, 0.3, 0.6, 0.95. Match the correlations with the data sets; two will be left over. Explain your answers.

1. Studies find a negative correlation between hours spent watching television and scores on reading tests. Does watching television make people less able to read? Discuss briefly.