Week 8: R

Griffin Dietz
CS 197 HCI Section
Draft paper assignment: 2-phase submission

MTurk Sandbox clarification and 3 stage deployment process

R! (I will circulate to talk to teams about evaluation plans)
Packages and Libraries

install.packages("dplyr") to install the first time you use a package, e.g., dplyr

library("dplyr") to load every time you use that package

For this tutorial, you will need tidyverse, which is a wrapper around a lot of packages useful for statistical analysis
Assignment operators: <- vs =

<- is an assignment operator in R

= is both an assignment operator and is used to specify function arguments

Example
plot <- ggplot(data=data, mapping=aes(x = age_group, y = correct)) +
  geom_line()

vs.

plot = ggplot(data=data, mapping=aes(x = age_group, y = correct)) +
  geom_line()

For clarity of code: use <-
# Tidy Data

One observation per row

Why? Consider: \( t\text{.test(cases~year)} \)

<table>
<thead>
<tr>
<th>State</th>
<th>1997</th>
<th>2002</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>101</td>
<td>142</td>
<td>165</td>
</tr>
<tr>
<td>AK</td>
<td>62</td>
<td>87</td>
<td>99</td>
</tr>
<tr>
<td>AZ</td>
<td>214</td>
<td>233</td>
<td>249</td>
</tr>
<tr>
<td>AR</td>
<td>31</td>
<td>39</td>
<td>52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>State</th>
<th>Year</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>AL</td>
<td>1997</td>
<td>101</td>
</tr>
<tr>
<td>AL</td>
<td>2002</td>
<td>142</td>
</tr>
<tr>
<td>AL</td>
<td>2007</td>
<td>165</td>
</tr>
<tr>
<td>AK</td>
<td>1997</td>
<td>62</td>
</tr>
<tr>
<td>AK</td>
<td>2002</td>
<td>87</td>
</tr>
</tbody>
</table>
Piping in R

String together nested functions with %>% (in dplyr package). As operations get more complex, piping helps make code more readable

round(sum(numbers), 2) is the same as…

numbers %>%
  sum() %>%
  round(2)  #piping passes in prior computation as first argument

Helpful tip: command-shift-m types a %>% in RStudio (control-shift-m on Windows)
https://github.com/gdietz44/R-tutorial