Lab Worksheet 4

Part 1: The dplyr pipe

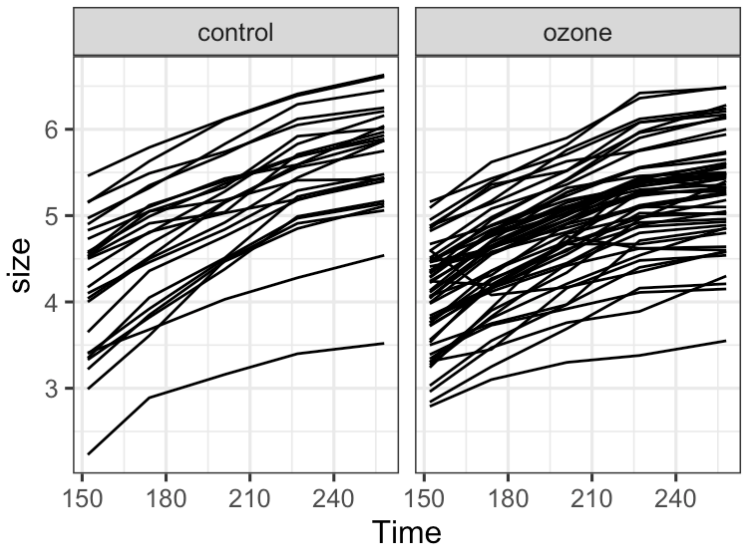
The following questions in Part 1 are from Lab Worksheet 3. Answer these questions again, but this time use the dplyr pipe (`%>%`) in your answer.

Problem 1: *In an in-class exercise, we made the following plot of the Sitka dataset:*

```
# download the sitka data set:
sitka <- read.csv("http://wilkelab.org/classes/SDS348/data_sets/sitka.csv")
head(sitka)
```

```
##   size Time tree treat
## 1 4.51  152    1 ozone
## 2 4.98  174    1 ozone
## 3 5.41  201    1 ozone
## 4 5.90  227    1 ozone
## 5 6.15  258    1 ozone
## 6 4.24  152    2 ozone
```

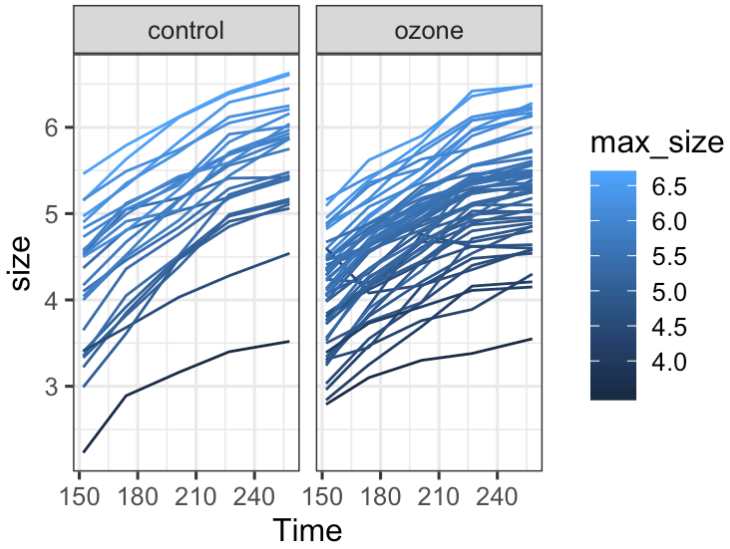
```
ggplot(sitka, aes(x = Time, y = size, group = tree)) +
  geom_line() +
  facet_wrap(~treat)
```



Now modify the plot so that the line for each tree is colored according to the maximum size of the tree.

```
sitka_new <-
  sitka %>%
  group_by(tree) %>%
  mutate(max_size = max(size))

ggplot(sitka_new, aes(x = Time, y = size, group = tree, color = max_size)) +
  geom_line() +
  facet_wrap(~treat)
```



Problem 2: *The package nycflights13 contains information about all flights departing from one of the NY City airports in 2013. In particular, the data table* `flights` *lists on-time departure and arrival information for 336,776 individual flights:*

```
library(nycflights13)
flights
```

```
## # A tibble: 336,776 x 19
##     year month   day dep_time sched_dep_time dep_delay arr_time
##    <int> <int> <int>    <int>          <int>     <dbl>    <int>
##  1  2013     1     1      517            515         2      830
##  2  2013     1     1      533            529         4      850
##  3  2013     1     1      542            540         2      923
##  4  2013     1     1      544            545        -1     1004
##  5  2013     1     1      554            600        -6      812
##  6  2013     1     1      554            558        -4      740
##  7  2013     1     1      555            600        -5      913
##  8  2013     1     1      557            600        -3      709
##  9  2013     1     1      557            600        -3      838
## 10  2013     1     1      558            600        -2      753
## # ... with 336,766 more rows, and 12 more variables: sched_arr_time <int>,
## #   arr_delay <dbl>, carrier <chr>, flight <int>, tailnum <chr>,
## #   origin <chr>, dest <chr>, air_time <dbl>, distance <dbl>, hour <dbl>,
## #   minute <dbl>, time_hour <dttm>
```

We would like to collect some information about arrival delays of United Airlines (UA) flights. Do the following: pick all UA departures with non-zero arrival delay and calculate the mean arrival delay for each of the corresponding flight numbers. Which flight had the longest mean arrival delay and how long was that delay?

```
flights %>%
  filter(carrier == "UA" & arr_delay != 0) %>%
  group_by(flight) %>%
  summarize(mean_delay = mean(arr_delay)) %>%
  filter(mean_delay == max(mean_delay))
```

```
## # A tibble: 1 x 2
##   flight mean_delay
##    <int>      <dbl>
## 1   1510        283
```

Flight 1510 had the longest delay, with an average arrival delay of 283 minutes.

###Part 2: Combining data-frames with dplyr

Problem 1: *Invent two simple data sets that allow you explain the difference between the dplyr functions* `left_join()` *and* `inner_join()`*. Explain which features of your data sets affect the behavior of these two functions.*

```
# data set 1
d1 <- read.table(text = "
label number1
A 1
B 2
C 4", head = T)

# data set 2
d2 <- read.table(text = "
label number2
A 2
C 4
D 6
", head = T)

d1
```

```
##   label number1
## 1     A       1
## 2     B       2
## 3     C       4
```

```
d2
```

```
##   label number2
## 1     A       2
## 2     C       4
## 3     D       6
```

```
left_join(d1, d2)
```

```
## Joining, by = "label"
```

```
## Warning: Column `label` joining factors with different levels, coercing to
## character vector
```

```
##   label number1 number2
## 1     A       1       2
## 2     B       2      NA
## 3     C       4       4
```

```
inner_join(d1, d2)
```

```
## Joining, by = "label"
```

```
## Warning: Column `label` joining factors with different levels, coercing to
## character vector
```

```
##   label number1 number2
## 1     A       1       2
## 2     C       4       4
```

We join the two data sets by `label`. `left_join()` finds all rows in the second data set that match to rows in the first data set. Therefore, we don't get a value for `number2` corresponding to label B. By contrast, `inner_join()` only keeps the matching rows that exist in both data sets. Therefore, the resulting table has only two rows, one for label A and one for label C.

Problem 2: *I have split the sitka data set into two data-frames. First, look up the documentation for the `bind_rows` function. What does `bind_rows` do? Next, use `bind_rows` to combine `sitka1` and `sitka2` back into a single data-frame.*

The `bind_rows` function adds rows from one data-frame to another as long as both data-frames have the same number of columns and identical column names.

```
sitka1 <- sitka[1:100, ]
sitka2 <- sitka[101:395, ]
head(sitka1)
```

```
##   size Time tree treat
## 1 4.51  152    1 ozone
## 2 4.98  174    1 ozone
## 3 5.41  201    1 ozone
## 4 5.90  227    1 ozone
## 5 6.15  258    1 ozone
## 6 4.24  152    2 ozone
```

```
head(sitka2)
```

```
##     size Time tree treat
## 101 4.04  152   21 ozone
## 102 4.64  174   21 ozone
## 103 4.86  201   21 ozone
## 104 5.09  227   21 ozone
## 105 5.25  258   21 ozone
## 106 3.53  152   22 ozone
```

```
sitka_combined <- bind_rows(sitka1, sitka2)
sitka_combined
```

```
##     size Time tree   treat
## 1   4.51  152    1   ozone
## 2   4.98  174    1   ozone
## 3   5.41  201    1   ozone
## 4   5.90  227    1   ozone
## 5   6.15  258    1   ozone
## 6   4.24  152    2   ozone
## 7   4.20  174    2   ozone
## 8   4.68  201    2   ozone
## 9   4.92  227    2   ozone
## 10  4.96  258    2   ozone
## 11  3.98  152    3   ozone
## 12  4.36  174    3   ozone
## 13  4.79  201    3   ozone
## 14  4.99  227    3   ozone
## 15  5.03  258    3   ozone
## 16  4.36  152    4   ozone
## 17  4.77  174    4   ozone
## 18  5.10  201    4   ozone
## 19  5.30  227    4   ozone
## 20  5.36  258    4   ozone
## 21  4.34  152    5   ozone
## 22  4.95  174    5   ozone
## 23  5.42  201    5   ozone
## 24  5.97  227    5   ozone
## 25  6.28  258    5   ozone
## 26  4.59  152    6   ozone
## 27  5.08  174    6   ozone
## 28  5.36  201    6   ozone
## 29  5.76  227    6   ozone
## 30  6.00  258    6   ozone
## 31  4.41  152    7   ozone
## 32  4.56  174    7   ozone
## 33  4.95  201    7   ozone
## 34  5.23  227    7   ozone
## 35  5.33  258    7   ozone
## 36  4.24  152    8   ozone
## 37  4.64  174    8   ozone
## 38  4.95  201    8   ozone
## 39  5.38  227    8   ozone
## 40  5.48  258    8   ozone
## 41  4.82  152    9   ozone
## 42  5.17  174    9   ozone
## 43  5.76  201    9   ozone
## 44  6.12  227    9   ozone
## 45  6.24  258    9   ozone
```

```
## 46  3.84  152   10   ozone
## 47  4.17  174   10   ozone
## 48  4.67  201   10   ozone
## 49  4.67  227   10   ozone
## 50  4.80  258   10   ozone
## 51  4.07  152   11   ozone
## 52  4.31  174   11   ozone
## 53  4.90  201   11   ozone
## 54  5.10  227   11   ozone
## 55  5.10  258   11   ozone
## 56  4.28  152   12   ozone
## 57  4.80  174   12   ozone
## 58  5.27  201   12   ozone
## 59  5.55  227   12   ozone
## 60  5.65  258   12   ozone
## 61  4.47  152   13   ozone
## 62  4.89  174   13   ozone
## 63  5.23  201   13   ozone
## 64  5.55  227   13   ozone
## 65  5.74  258   13   ozone
## 66  4.46  152   14   ozone
## 67  4.84  174   14   ozone
## 68  5.11  201   14   ozone
## 69  5.34  227   14   ozone
## 70  5.46  258   14   ozone
## 71  4.60  152   15   ozone
## 72  4.08  174   15   ozone
## 73  4.17  201   15   ozone
## 74  4.35  227   15   ozone
## 75  4.59  258   15   ozone
## 76  3.73  152   16   ozone
## 77  4.15  174   16   ozone
## 78  4.61  201   16   ozone
## 79  4.87  227   16   ozone
## 80  4.93  258   16   ozone
## 81  4.67  152   17   ozone
## 82  4.88  174   17   ozone
## 83  5.18  201   17   ozone
## 84  5.34  227   17   ozone
## 85  5.49  258   17   ozone
## 86  2.96  152   18   ozone
## 87  3.47  174   18   ozone
## 88  3.76  201   18   ozone
## 89  3.89  227   18   ozone
## 90  4.30  258   18   ozone
## 91  3.24  152   19   ozone
## 92  3.93  174   19   ozone
```

```
## 93  4.76  201   19   ozone
## 94  4.62  227   19   ozone
## 95  4.64  258   19   ozone
## 96  4.36  152   20   ozone
## 97  4.77  174   20   ozone
## 98  5.02  201   20   ozone
## 99  5.26  227   20   ozone
## 100 5.45  258   20   ozone
## 101 4.04  152   21   ozone
## 102 4.64  174   21   ozone
## 103 4.86  201   21   ozone
## 104 5.09  227   21   ozone
## 105 5.25  258   21   ozone
## 106 3.53  152   22   ozone
## 107 4.25  174   22   ozone
## 108 4.68  201   22   ozone
## 109 4.97  227   22   ozone
## 110 5.18  258   22   ozone
## 111 4.22  152   23   ozone
## 112 4.69  174   23   ozone
## 113 5.07  201   23   ozone
## 114 5.37  227   23   ozone
## 115 5.58  258   23   ozone
## 116 2.79  152   24   ozone
## 117 3.10  174   24   ozone
## 118 3.30  201   24   ozone
## 119 3.38  227   24   ozone
## 120 3.55  258   24   ozone
## 121 3.30  152   25   ozone
## 122 3.90  174   25   ozone
## 123 4.34  201   25   ozone
## 124 4.96  227   25   ozone
## 125 5.40  258   25   ozone
## 126 3.34  152   26   ozone
## 127 3.81  174   26   ozone
## 128 4.21  201   26   ozone
## 129 4.54  227   26   ozone
## 130 4.86  258   26   ozone
## 131 3.76  152   27   ozone
## 132 4.36  174   27   ozone
## 133 4.70  201   27   ozone
## 134 5.44  227   27   ozone
## 135 5.32  258   27   ozone
## 136 4.49  152   28   ozone
## 137 4.76  174   28   ozone
## 138 5.15  201   28   ozone
## 139 5.37  227   28   ozone
```

```
## 140 5.56  258   28   ozone
## 141 4.88  152   29   ozone
## 142 5.14  174   29   ozone
## 143 5.52  201   29   ozone
## 144 6.08  227   29   ozone
## 145 6.17  258   29   ozone
## 146 4.88  152   30   ozone
## 147 5.32  174   30   ozone
## 148 5.63  201   30   ozone
## 149 5.75  227   30   ozone
## 150 5.94  258   30   ozone
## 151 3.80  152   31   ozone
## 152 4.16  174   31   ozone
## 153 4.45  201   31   ozone
## 154 4.89  227   31   ozone
## 155 5.05  258   31   ozone
## 156 4.46  152   32   ozone
## 157 4.62  174   32   ozone
## 158 5.00  201   32   ozone
## 159 5.40  227   32   ozone
## 160 5.49  258   32   ozone
## 161 4.29  152   33   ozone
## 162 4.82  174   33   ozone
## 163 5.32  201   33   ozone
## 164 5.46  227   33   ozone
## 165 5.50  258   33   ozone
## 166 4.06  152   34   ozone
## 167 4.58  174   34   ozone
## 168 4.81  201   34   ozone
## 169 5.12  227   34   ozone
## 170 5.27  258   34   ozone
## 171 5.16  152   35   ozone
## 172 5.43  174   35   ozone
## 173 5.71  201   35   ozone
## 174 6.08  227   35   ozone
## 175 6.21  258   35   ozone
## 176 3.81  152   36   ozone
## 177 4.12  174   36   ozone
## 178 4.42  201   36   ozone
## 179 4.62  227   36   ozone
## 180 4.60  258   36   ozone
## 181 5.09  152   37   ozone
## 182 5.62  174   37   ozone
## 183 5.90  201   37   ozone
## 184 6.36  227   37   ozone
## 185 6.49  258   37   ozone
## 186 4.13  152   38   ozone
```

```
## 187 4.71  174   38   ozone
## 188 5.27  201   38   ozone
## 189 5.56  227   38   ozone
## 190 5.72  258   38   ozone
## 191 4.85  152   39   ozone
## 192 5.36  174   39   ozone
## 193 5.52  201   39   ozone
## 194 5.96  227   39   ozone
## 195 6.13  258   39   ozone
## 196 4.11  152   40   ozone
## 197 4.62  174   40   ozone
## 198 4.95  201   40   ozone
## 199 5.28  227   40   ozone
## 200 5.43  258   40   ozone
## 201 4.95  152   41   ozone
## 202 5.39  174   41   ozone
## 203 5.82  201   41   ozone
## 204 6.42  227   41   ozone
## 205 6.48  258   41   ozone
## 206 4.36  152   42   ozone
## 207 4.65  174   42   ozone
## 208 5.04  201   42   ozone
## 209 5.38  227   42   ozone
## 210 5.47  258   42   ozone
## 211 4.05  152   43   ozone
## 212 4.65  174   43   ozone
## 213 5.09  201   43   ozone
## 214 5.44  227   43   ozone
## 215 5.60  258   43   ozone
## 216 3.76  152   44   ozone
## 217 4.27  174   44   ozone
## 218 4.59  201   44   ozone
## 219 5.10  227   44   ozone
## 220 5.25  258   44   ozone
## 221 2.84  152   45   ozone
## 222 3.25  174   45   ozone
## 223 3.69  201   45   ozone
## 224 4.16  227   45   ozone
## 225 4.21  258   45   ozone
## 226 4.33  152   46   ozone
## 227 4.80  174   46   ozone
## 228 5.09  201   46   ozone
## 229 5.42  227   46   ozone
## 230 5.61  258   46   ozone
## 231 3.99  152   47   ozone
## 232 4.55  174   47   ozone
## 233 4.91  201   47   ozone
```

```
## 234 5.26  227   47   ozone
## 235 5.30  258   47   ozone
## 236 3.50  152   48   ozone
## 237 3.75  174   48   ozone
## 238 3.97  201   48   ozone
## 239 4.71  227   48   ozone
## 240 4.85  258   48   ozone
## 241 3.31  152   49   ozone
## 242 3.45  174   49   ozone
## 243 4.16  201   49   ozone
## 244 4.48  227   49   ozone
## 245 4.54  258   49   ozone
## 246 3.03  152   50   ozone
## 247 3.55  174   50   ozone
## 248 3.97  201   50   ozone
## 249 4.40  227   50   ozone
## 250 4.58  258   50   ozone
## 251 3.27  152   51   ozone
## 252 3.83  174   51   ozone
## 253 4.44  201   51   ozone
## 254 4.80  227   51   ozone
## 255 4.89  258   51   ozone
## 256 3.56  152   52   ozone
## 257 4.18  174   52   ozone
## 258 4.70  201   52   ozone
## 259 5.27  227   52   ozone
## 260 5.28  258   52   ozone
## 261 3.39  152   53   ozone
## 262 3.73  174   53   ozone
## 263 3.92  201   53   ozone
## 264 4.11  227   53   ozone
## 265 4.15  258   53   ozone
## 266 3.72  152   54   ozone
## 267 4.16  174   54   ozone
## 268 4.55  201   54   ozone
## 269 5.03  227   54   ozone
## 270 5.02  258   54   ozone
## 271 4.53  152   55 control
## 272 5.05  174   55 control
## 273 5.18  201   55 control
## 274 5.41  227   55 control
## 275 5.42  258   55 control
## 276 4.97  152   56 control
## 277 5.32  174   56 control
## 278 5.83  201   56 control
## 279 6.29  227   56 control
## 280 6.45  258   56 control
```

```
## 281 4.37  152   57 control
## 282 4.81  174   57 control
## 283 5.03  201   57 control
## 284 5.19  227   57 control
## 285 5.40  258   57 control
## 286 4.58  152   58 control
## 287 4.99  174   58 control
## 288 5.37  201   58 control
## 289 5.68  227   58 control
## 290 5.93  258   58 control
## 291 4.00  152   59 control
## 292 4.50  174   59 control
## 293 4.92  201   59 control
## 294 5.44  227   59 control
## 295 5.87  258   59 control
## 296 4.73  152   60 control
## 297 5.05  174   60 control
## 298 5.33  201   60 control
## 299 5.92  227   60 control
## 300 6.01  258   60 control
## 301 5.15  152   61 control
## 302 5.63  174   61 control
## 303 6.11  201   61 control
## 304 6.39  227   61 control
## 305 6.61  258   61 control
## 306 4.10  152   62 control
## 307 4.46  174   62 control
## 308 4.84  201   62 control
## 309 5.29  227   62 control
## 310 5.48  258   62 control
## 311 3.22  152   63 control
## 312 3.85  174   63 control
## 313 4.47  201   63 control
## 314 4.85  227   63 control
## 315 5.11  258   63 control
## 316 2.23  152   64 control
## 317 2.89  174   64 control
## 318 3.16  201   64 control
## 319 3.40  227   64 control
## 320 3.52  258   64 control
## 321 3.65  152   65 control
## 322 4.36  174   65 control
## 323 4.76  201   65 control
## 324 5.18  227   65 control
## 325 5.44  258   65 control
## 326 3.40  152   66 control
## 327 3.92  174   66 control
```

```
## 328 4.50  201   66 control
## 329 4.97  227   66 control
## 330 5.14  258   66 control
## 331 5.16  152   67 control
## 332 5.49  174   67 control
## 333 5.74  201   67 control
## 334 6.05  227   67 control
## 335 6.21  258   67 control
## 336 4.04  152   68 control
## 337 4.52  174   68 control
## 338 5.15  201   68 control
## 339 5.59  227   68 control
## 340 5.87  258   68 control
## 341 4.52  152   69 control
## 342 4.91  174   69 control
## 343 5.04  201   69 control
## 344 5.71  227   69 control
## 345 5.97  258   69 control
## 346 4.56  152   70 control
## 347 5.12  174   70 control
## 348 5.40  201   70 control
## 349 5.69  227   70 control
## 350 5.89  258   70 control
## 351 4.90  152   71 control
## 352 5.35  174   71 control
## 353 5.71  201   71 control
## 354 6.12  227   71 control
## 355 6.25  258   71 control
## 356 4.83  152   72 control
## 357 5.10  174   72 control
## 358 5.43  201   72 control
## 359 5.59  227   72 control
## 360 6.04  258   72 control
## 361 5.46  152   73 control
## 362 5.79  174   73 control
## 363 6.12  201   73 control
## 364 6.41  227   73 control
## 365 6.63  258   73 control
## 366 4.17  152   74 control
## 367 4.67  174   74 control
## 368 5.16  201   74 control
## 369 5.56  227   74 control
## 370 5.75  258   74 control
## 371 3.35  152   75 control
## 372 4.05  174   75 control
## 373 4.51  201   75 control
## 374 5.22  227   75 control
```

```
## 375 5.44  258   75 control
## 376 3.33  152   76 control
## 377 3.82  174   76 control
## 378 4.38  201   76 control
## 379 4.99  227   76 control
## 380 5.17  258   76 control
## 381 3.41  152   77 control
## 382 3.68  174   77 control
## 383 4.03  201   77 control
## 384 4.28  227   77 control
## 385 4.54  258   77 control
## 386 4.50  152   78 control
## 387 4.80  174   78 control
## 388 5.28  201   78 control
## 389 5.83  227   78 control
## 390 6.16  258   78 control
## 391 2.99  152   79 control
## 392 3.61  174   79 control
## 393 4.48  201   79 control
## 394 4.91  227   79 control
## 395 5.06  258   79 control
```