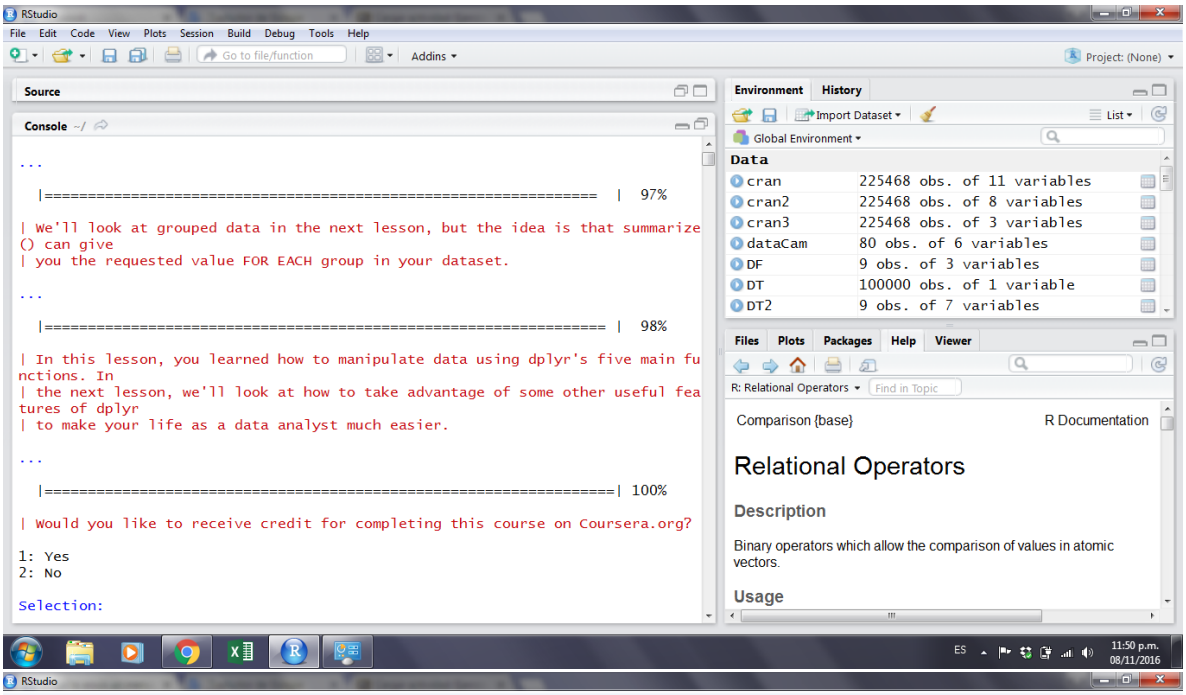
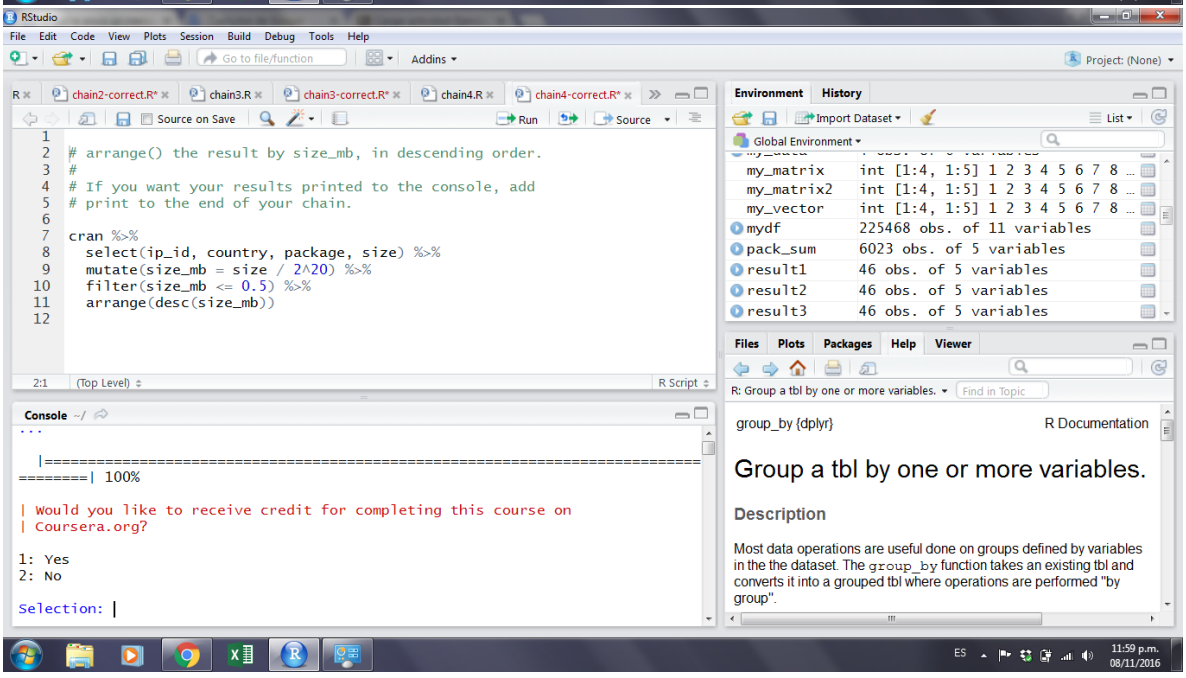
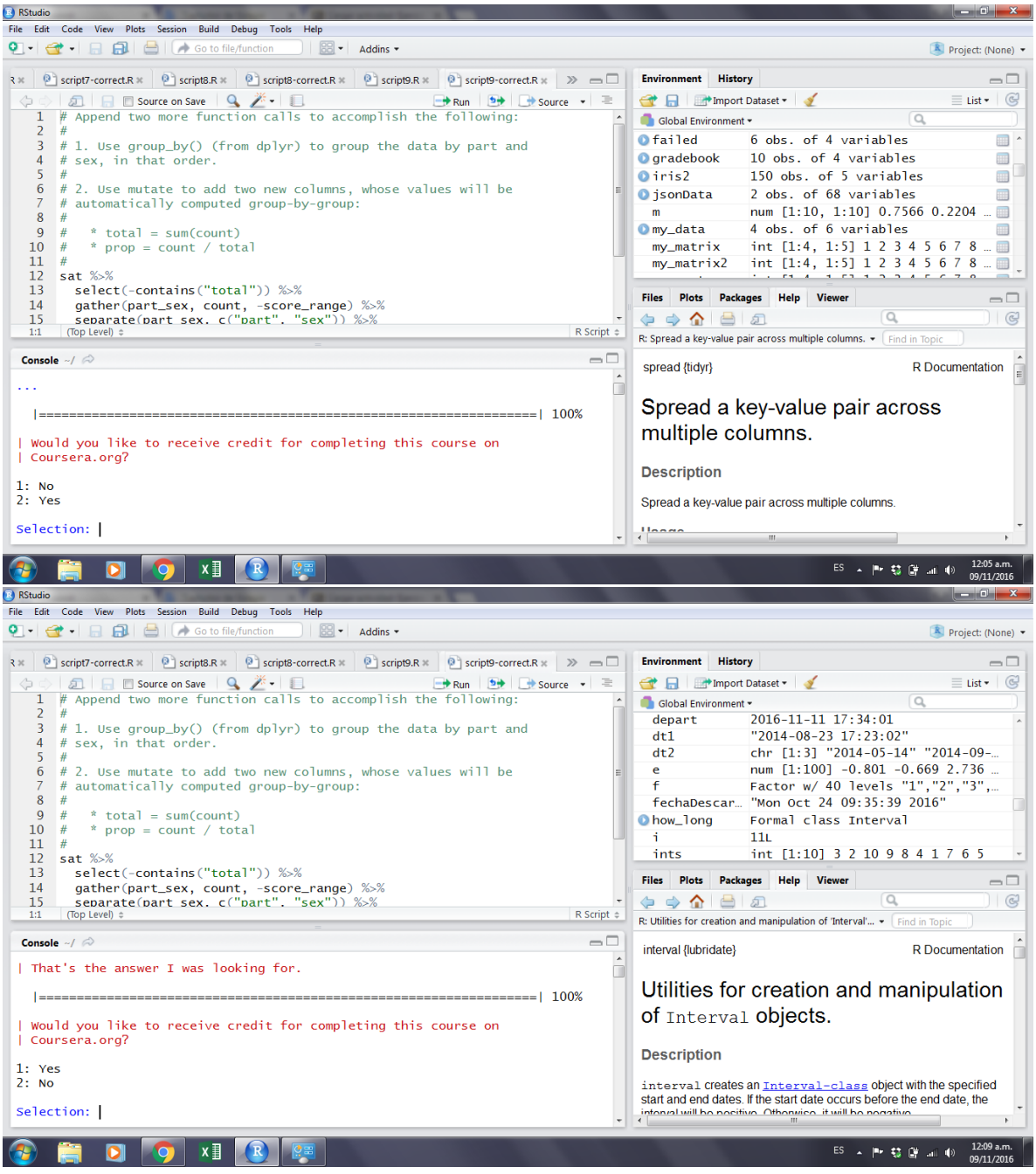

RStudio
File Edit Code View Plots Session Build Debug Tools Help
Source
Console ~/
...
|========== | 97%
| We'll look at grouped data in the next lesson, but the idea is that summarize() can give
| you the requested value FOR EACH group in your dataset.
...
|========== | 98%
| In this lesson, you learned how to manipulate data using dplyr's five main functions. In
| the next lesson, we'll look at how to take advantage of some other useful features of dplyr
| to make your life as a data analyst much easier.
...
|==========| 100%
| Would you like to receive credit for completing this course on Coursera.org?
1: Yes
2: No
Selection:
Environment History
Import Dataset
Global Environment
Data
cran 225468 obs. of 11 variables
cran2 225468 obs. of 8 variables
cran3 225468 obs. of 3 variables
dataCam 80 obs. of 6 variables
DF 9 obs. of 3 variables
DT 100000 obs. of 1 variable
DT2 9 obs. of 7 variables
Files Plots Packages Help Viewer
R: Relational Operators
Comparison {base} R Documentation
Relational Operators
Description
Binary operators which allow the comparison of values in atomic vectors.
Usage


RStudio
File Edit Code View Plots Session Build Debug Tools Help
chain2-correct.R* chain3.R chain3-correct.R* chain4.R chain4-correct.R*
Source on Save Run Source
arrange() the result by size_mb, in descending order.
#
If you want your results printed to the console, add
print to the end of your chain.
cran %>%
select(ip_id, country, package, size) %>%
mutate(size_mb = size / 2^20) %>%
filter(size_mb <= 0.5) %>%
arrange(desc(size_mb))
2:1 (Top Level) R Script
Console ~/
...
|==========| 100%
| Would you like to receive credit for completing this course on Coursera.org?
1: Yes
2: No
Selection:
Environment History
Import Dataset
Global Environment
my_matrix int [1:4, 1:5] 1 2 3 4 5 6 7 8 ...
my_matrix2 int [1:4, 1:5] 1 2 3 4 5 6 7 8 ...
my_vector int [1:4, 1:5] 1 2 3 4 5 6 7 8 ...
mydf 225468 obs. of 11 variables
pack_sum 6023 obs. of 5 variables
result1 46 obs. of 5 variables
result2 46 obs. of 5 variables
result3 46 obs. of 5 variables
Files Plots Packages Help Viewer
R: Group a tbl by one or more variables.
group_by {dplyr} R Documentation
Group a tbl by one or more variables.
Description
Most data operations are useful done on groups defined by variables in the the dataset. The group_by function takes an existing tbl and converts it into a grouped tbl where operations are performed "by group".


RStudio
File Edit Code View Plots Session Build Debug Tools Help
Addins
Project: (None)
script7-correct.R
script8.R
script8-correct.R
script9.R
script9-correct.R
Source on Save
Run
Source
Append two more function calls to accomplish the following:
#
1. Use group_by() (from dplyr) to group the data by part and
sex, in that order.
#
2. Use mutate to add two new columns, whose values will be
automatically computed group-by-group:
#
* total = sum(count)
* prop = count / total
#
sat %>%
  select(-contains("total")) %>%
  gather(part_sex, count, -score_range) %>%
  separate(part_sex, c("part", "sex")) %>%
(Top Level)
R Script
Environment
History
Import Dataset
List
Global Environment
failed 6 obs. of 4 variables
gradebook 10 obs. of 4 variables
iris2 150 obs. of 5 variables
jsonData 2 obs. of 68 variables
m num [1:10, 1:10] 0.7566 0.2204 ...
my_data 4 obs. of 6 variables
my_matrix int [1:4, 1:5] 1 2 3 4 5 6 7 8 ...
my_matrix2 int [1:4, 1:5] 1 2 3 4 5 6 7 8 ...
Files
Plots
Packages
Help
Viewer
R: Spread a key-value pair across multiple columns.
Find in Topic
Console ~/
...
|==| 100%
| Would you like to receive credit for completing this course on
| Coursera.org?
1: No
2: Yes
Selection:
spread {tidyr}
R Documentation
Spread a key-value pair across multiple columns.
Description
Spread a key-value pair across multiple columns.
ES
12:05 a.m.
09/11/2016
RStudio
File Edit Code View Plots Session Build Debug Tools Help
Addins
Project: (None)
script7-correct.R
script8.R
script8-correct.R
script9.R
script9-correct.R
Source on Save
Run
Source
Append two more function calls to accomplish the following:
#
1. Use group_by() (from dplyr) to group the data by part and
sex, in that order.
#
2. Use mutate to add two new columns, whose values will be
automatically computed group-by-group:
#
* total = sum(count)
* prop = count / total
#
sat %>%
  select(-contains("total")) %>%
  gather(part_sex, count, -score_range) %>%
  separate(part_sex, c("part", "sex")) %>%
(Top Level)
R Script
Environment
History
Import Dataset
List
Global Environment
depart 2016-11-11 17:34:01
dt1 "2014-08-23 17:23:02"
dt2 chr [1:3] "2014-05-14" "2014-09-...
e num [1:100] -0.801 -0.669 2.736 ...
f Factor w/ 40 levels "1","2","3",...
fechaDescar... "Mon Oct 24 09:35:39 2016"
how_long Formal class Interval
i 11L
ints int [1:10] 3 2 10 9 8 4 1 7 6 5
Files
Plots
Packages
Help
Viewer
R: Utilities for creation and manipulation of 'Interval'...
Find in Topic
Console ~/
| That's the answer I was looking for.
|==| 100%
| Would you like to receive credit for completing this course on
| Coursera.org?
1: Yes
2: No
Selection:
interval {lubridate}
R Documentation
Utilities for creation and manipulation of Interval objects.
Description
interval creates an Interval-class object with the specified
ES
12:09 a.m.
09/11/2016