# dplyr package

## Merge, rename, select, filter

library(dplyr)

data2 <- read.csv("data2.csv")

data1 <- read.csv("data1.csv")

mergedData <- merge(data1, data2, by.x = "X", by.y = "CountryCode")

mergedData <- rename(mergedData, CountryCode = X, Country = X.2, IncomeGroup=Income.Group, GDP = Gross.domestic.product.2012)

mergedData <- select(mergedData, CountryCode, Country, IncomeGroup, GDP)

mergedData[,4] <- as.numeric(as.character(mergedData[,4]))

mergedData <- filter(mergedData, GDP>0)

## Grouping by 5 equal parts, table

mergedData$group <- cut2(mergedData$GDP, g=5)

table(mergedData$group, mergedData$IncomeGroup)

## Grouping by extracting top 1% of variable

quantile(pack\_sum$count, probs = 0.99)

top\_counts <- filter(pack\_sum, count > 679)

top\_counts\_sorted <- arrange(top\_counts, desc(count))

## Group by variable, summarize, pipeline operator

result3 <- cran %>%

group\_by(package) %>%

summarize(count = n(),

unique = n\_distinct(ip\_id),

countries = n\_distinct(country),

avg\_bytes = mean(size)

) %>%

filter(countries > 60) %>%

arrange(desc(countries), avg\_bytes)

# tidyr package

> students2

grade male\_1 female\_1 male\_2 female\_2

A 3 4 3 4

B 6 4 3 5

C 7 4 3 8

D 4 0 8 1

E 1 1 2 7

students2 %>%

gather(sex\_class,count,-grade) %>%

separate(sex\_class, c("sex", "class")) %>%

print

sat %>%

select(-contains("total")) %>%

gather(part\_sex, count, -score\_range) %>%

separate(part\_sex, c("part", "sex")) %>%

print