Unite Examples

# Some Examples of spread, gather, unite, separate.

See the [R Studio Data Wrangling Cheatsheet](https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf).

See the [R Studio ggplot Cheatsheet](https://www.rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf).

library(tidyverse)

### The examples from the cheatsheet.

mydata <- data\_frame(  
 a = 1:3,   
 b = 4:6  
)  
  
mydata

## # A tibble: 3 x 2  
## a b  
## <int> <int>  
## 1 1 4  
## 2 2 5  
## 3 3 6

Arrange the data.

arrange(mydata, desc(a))

## # A tibble: 3 x 2  
## a b  
## <int> <int>  
## 1 3 6  
## 2 2 5  
## 3 1 4

Rename the data.

mydata <- rename(mydata, x = a, y = b)  
  
mydata

## # A tibble: 3 x 2  
## x y  
## <int> <int>  
## 1 1 4  
## 2 2 5  
## 3 3 6

arrange(mydata, desc(x))

## # A tibble: 3 x 2  
## x y  
## <int> <int>  
## 1 3 6  
## 2 2 5  
## 3 1 4

mydata %>% mutate(x.prop = x/sum(x), x.cum.prop = cumsum(x)/sum(x), y.prop = y/sum(y), y.cum.prop = cumsum(y)/sum(y)) %>%  
 select(x, x.prop, x.cum.prop, y, y.prop, y.cum.prop)

## # A tibble: 3 x 6  
## x x.prop x.cum.prop y y.prop y.cum.prop  
## <int> <dbl> <dbl> <int> <dbl> <dbl>  
## 1 1 0.167 0.167 4 0.267 0.267  
## 2 2 0.333 0.5 5 0.333 0.6   
## 3 3 0.5 1 6 0.4 1

# Example, page 27, Problem 2.2, Ott 3rd Edition

imports <- data\_frame(  
 Year = c(1979:1986),  
 Import = c(17518,15491,19898,16663,17061,26171,23650,19650)  
)  
  
imports

## # A tibble: 8 x 2  
## Year Import  
## <int> <dbl>  
## 1 1979 17518  
## 2 1980 15491  
## 3 1981 19898  
## 4 1982 16663  
## 5 1983 17061  
## 6 1984 26171  
## 7 1985 23650  
## 8 1986 19650

imports.wide <- spread(imports, key = Year, value = Import)  
imports.wide

## # A tibble: 1 x 8  
## `1979` `1980` `1981` `1982` `1983` `1984` `1985` `1986`  
## <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 17518 15491 19898 16663 17061 26171 23650 19650

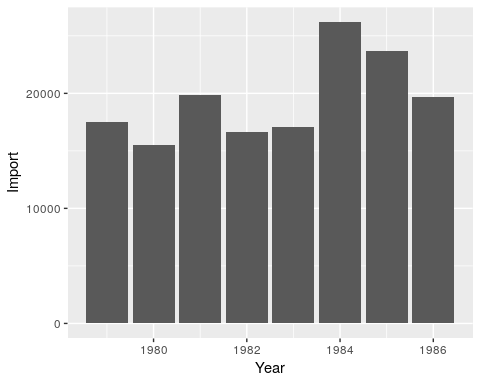
imports.narrow <- gather(imports.wide, key = "Year.New", value = Import.New)  
imports.narrow

## # A tibble: 8 x 2  
## Year.New Import.New  
## <chr> <dbl>  
## 1 1979 17518  
## 2 1980 15491  
## 3 1981 19898  
## 4 1982 16663  
## 5 1983 17061  
## 6 1984 26171  
## 7 1985 23650  
## 8 1986 19650

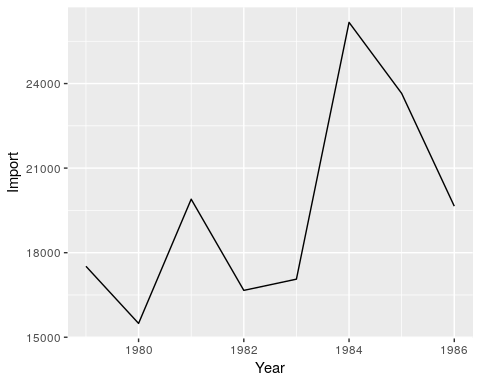
imports.narrow <- imports.narrow %>% mutate(Year.New = as.integer(Year.New))  
imports.narrow

## # A tibble: 8 x 2  
## Year.New Import.New  
## <int> <dbl>  
## 1 1979 17518  
## 2 1980 15491  
## 3 1981 19898  
## 4 1982 16663  
## 5 1983 17061  
## 6 1984 26171  
## 7 1985 23650  
## 8 1986 19650

imports %>% ggplot(aes(x=Year, y=Import)) +  
 geom\_col()



imports %>% ggplot(aes(x=Year, y=Import)) +  
 geom\_line()



# Example, page 28, Problem 2.4, Ott 3rd Edition

GNP.1985 <- data\_frame(  
 Year = c(1985,1985,1985,1985),  
 Quarter = c("I","II","III","IV"),  
 GNP = c(3910,3961,4017,4067),  
 DPI = c(2505,2532,2503,2533)  
)  
GNP.1985

## # A tibble: 4 x 4  
## Year Quarter GNP DPI  
## <dbl> <chr> <dbl> <dbl>  
## 1 1985 I 3910 2505  
## 2 1985 II 3961 2532  
## 3 1985 III 4017 2503  
## 4 1985 IV 4067 2533

GNP.1986 <- data\_frame(  
 Year = c(1986,1986,1986,1986),  
 Quarter = c("I","II","III","IV"),  
 GNP = c(4137,4203,4266,4308),  
 DPI = c(2536,2555,2579,2589)  
)  
GNP.1986

## # A tibble: 4 x 4  
## Year Quarter GNP DPI  
## <dbl> <chr> <dbl> <dbl>  
## 1 1986 I 4137 2536  
## 2 1986 II 4203 2555  
## 3 1986 III 4266 2579  
## 4 1986 IV 4308 2589

GNP <- bind\_rows(GNP.1985, GNP.1986)  
GNP

## # A tibble: 8 x 4  
## Year Quarter GNP DPI  
## <dbl> <chr> <dbl> <dbl>  
## 1 1985 I 3910 2505  
## 2 1985 II 3961 2532  
## 3 1985 III 4017 2503  
## 4 1985 IV 4067 2533  
## 5 1986 I 4137 2536  
## 6 1986 II 4203 2555  
## 7 1986 III 4266 2579  
## 8 1986 IV 4308 2589

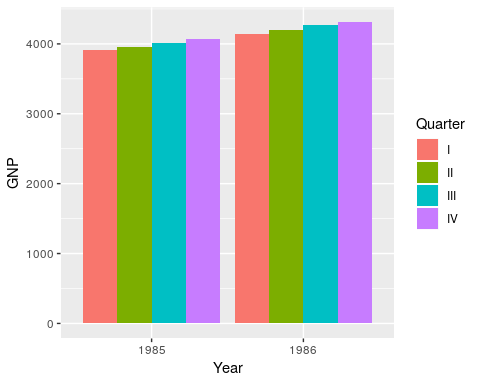
GNP.wide <- GNP %>% select(Year, Quarter, GNP) %>%  
 spread(key = Quarter, value = GNP )   
GNP.wide

## # A tibble: 2 x 5  
## Year I II III IV  
## <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 1985 3910 3961 4017 4067  
## 2 1986 4137 4203 4266 4308

GNP.narrow <- GNP.wide %>%  
 gather(key = Quarter, value = GNP, I,II,III, IV ) %>%  
 arrange(Year)  
GNP.narrow

## # A tibble: 8 x 3  
## Year Quarter GNP  
## <dbl> <chr> <dbl>  
## 1 1985 I 3910  
## 2 1985 II 3961  
## 3 1985 III 4017  
## 4 1985 IV 4067  
## 5 1986 I 4137  
## 6 1986 II 4203  
## 7 1986 III 4266  
## 8 1986 IV 4308

GNP %>% ggplot(aes(x= factor(Year), y = GNP, fill = Quarter)) +  
 geom\_bar(stat = "identity", position = "dodge") +  
 labs(x = "Year")



# Example, page 30, Problem 2.14, Ott 3rd Edition

SAT <- data\_frame(  
 GT = c("Male, Math", "Female, Math", "Male, Verbal", "Female, Verbal"),  
 "Year 1967" = c(514,467,463,486),  
 "Year 1970" = c(509,465,459,461),  
 "Year 1975" = c(495,449,437,431),  
 "Year 1980" = c(491,443,428,420),  
 "Year 1983" = c(493,445,430,420)  
)  
SAT

## # A tibble: 4 x 6  
## GT `Year 1967` `Year 1970` `Year 1975` `Year 1980` `Year 1983`  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Male, Math 514 509 495 491 493  
## 2 Female, Math 467 465 449 443 445  
## 3 Male, Verbal 463 459 437 428 430  
## 4 Female, Ver… 486 461 431 420 420

SAT.wide <- SAT %>% rename("1967" = "Year 1967", "1970" = "Year 1970", "1975" = "Year 1975", "1980" = "Year 1980", "1983" = "Year 1983")  
SAT.wide

## # A tibble: 4 x 6  
## GT `1967` `1970` `1975` `1980` `1983`  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Male, Math 514 509 495 491 493  
## 2 Female, Math 467 465 449 443 445  
## 3 Male, Verbal 463 459 437 428 430  
## 4 Female, Verbal 486 461 431 420 420

SAT.narrow <- SAT.wide %>% gather( key = Year, value = Score, "1967", "1970", "1975", "1980", "1983" )  
SAT.narrow

## # A tibble: 20 x 3  
## GT Year Score  
## <chr> <chr> <dbl>  
## 1 Male, Math 1967 514  
## 2 Female, Math 1967 467  
## 3 Male, Verbal 1967 463  
## 4 Female, Verbal 1967 486  
## 5 Male, Math 1970 509  
## 6 Female, Math 1970 465  
## 7 Male, Verbal 1970 459  
## 8 Female, Verbal 1970 461  
## 9 Male, Math 1975 495  
## 10 Female, Math 1975 449  
## 11 Male, Verbal 1975 437  
## 12 Female, Verbal 1975 431  
## 13 Male, Math 1980 491  
## 14 Female, Math 1980 443  
## 15 Male, Verbal 1980 428  
## 16 Female, Verbal 1980 420  
## 17 Male, Math 1983 493  
## 18 Female, Math 1983 445  
## 19 Male, Verbal 1983 430  
## 20 Female, Verbal 1983 420

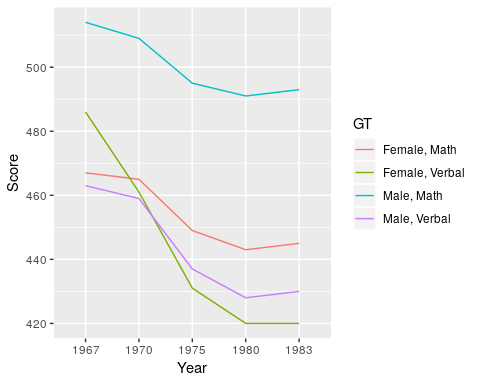
SAT.narrow2 <- SAT.narrow %>% separate(GT, c("Gender", "Type"))  
SAT.narrow2

## # A tibble: 20 x 4  
## Gender Type Year Score  
## <chr> <chr> <chr> <dbl>  
## 1 Male Math 1967 514  
## 2 Female Math 1967 467  
## 3 Male Verbal 1967 463  
## 4 Female Verbal 1967 486  
## 5 Male Math 1970 509  
## 6 Female Math 1970 465  
## 7 Male Verbal 1970 459  
## 8 Female Verbal 1970 461  
## 9 Male Math 1975 495  
## 10 Female Math 1975 449  
## 11 Male Verbal 1975 437  
## 12 Female Verbal 1975 431  
## 13 Male Math 1980 491  
## 14 Female Math 1980 443  
## 15 Male Verbal 1980 428  
## 16 Female Verbal 1980 420  
## 17 Male Math 1983 493  
## 18 Female Math 1983 445  
## 19 Male Verbal 1983 430  
## 20 Female Verbal 1983 420

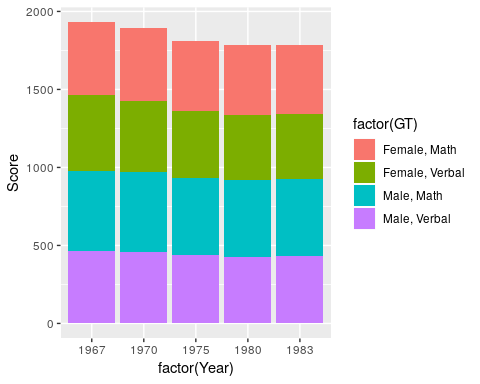
SAT.narrow3 <- SAT.narrow2 %>% unite(GT, c("Gender", "Type"), sep="\_")  
SAT.narrow3

## # A tibble: 20 x 3  
## GT Year Score  
## <chr> <chr> <dbl>  
## 1 Male\_Math 1967 514  
## 2 Female\_Math 1967 467  
## 3 Male\_Verbal 1967 463  
## 4 Female\_Verbal 1967 486  
## 5 Male\_Math 1970 509  
## 6 Female\_Math 1970 465  
## 7 Male\_Verbal 1970 459  
## 8 Female\_Verbal 1970 461  
## 9 Male\_Math 1975 495  
## 10 Female\_Math 1975 449  
## 11 Male\_Verbal 1975 437  
## 12 Female\_Verbal 1975 431  
## 13 Male\_Math 1980 491  
## 14 Female\_Math 1980 443  
## 15 Male\_Verbal 1980 428  
## 16 Female\_Verbal 1980 420  
## 17 Male\_Math 1983 493  
## 18 Female\_Math 1983 445  
## 19 Male\_Verbal 1983 430  
## 20 Female\_Verbal 1983 420

SAT.narrow %>% ggplot(aes(x=Year, y=Score, color=GT) ) +  
 geom\_line(aes(group = GT))



SAT.narrow %>% ggplot( aes( x=factor(Year), y=Score, fill=factor(GT) ) ) +  
 geom\_col()



SAT.narrow %>% ggplot( aes( x=factor(Year), y=Score, fill=factor(GT) ) ) +  
 geom\_col(position = "dodge")

