

Solucion Lab #1

Objetivos

1. Solucion del laboratorio de la union de tablas en un mismo directorio

Archivos de Texto

Cargando librerias

```
require(dplyr)
```

```
## Loading required package: dplyr
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
require(readxl)
```

```
## Loading required package: readxl
```

Funcion list.files()

Asignando la lista de archivos de la carpeta a una variable

```
files <- list.files(path = "C:/Users/JR29/Documents/GitHub/data-wrangling/data/01")  
files
```

```
## [1] "01-2017.xlsx" "02-2017.xlsx" "03-2017.xlsx" "04-2017.xlsx"
## [5] "05-2017.xlsx" "06-2017.xlsx" "07-2017.xlsx" "08-2017.xlsx"
## [9] "09-2017.xlsx" "10-2017.xlsx" "11-2017.xlsx"
```

Funcion read_excel()

Lectura del primer archivo de Excel de la carpeta 01

```
inv <- read_excel(files[1], #sheet = "sheet1"
                  col_types = "text")
str(inv)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   192 obs. of  8 variables:
## $ COD_VIAJE: chr  "10000001" "10000002" "10000003" "10000004" ...
## $ CLIENTE  : chr  "EL PINCHE OBELISCO / Despacho a cliente" "TAQUERIA EL CHINITO |
||Faltante" "TIENDA LA BENDICION / Despacho a cliente" "TAQUERIA EL CHINITO" ...
## $ UBICACION: chr  "76002" "76002" "76002" "76002" ...
## $ CANTIDAD : chr  "1200" "1433" "1857" "339" ...
## $ PILOTO   : chr  "Fernando Mariano Berrio" "Hector Aragonés Frutos" "Pedro Alvare
z Parejo" "Angel Valdez Alegria" ...
## $ Q        : chr  "300" "358.25" "464.25" "84.75" ...
## $ CREDITO  : chr  "30" "90" "60" "30" ...
## $ UNIDAD   : chr  "Camion Grande" "Camion Grande" "Camion Grande" "Panel" ...
```

Determinar el tipo de variable por columna

```
#inv$COD_VIAJE se queda en texto
#inv$CLIENTE se queda en texto
inv$UBICACION <- as.factor(inv$UBICACION)
inv$CANTIDAD <- as.numeric(inv$CANTIDAD)
#inv$PILOTO se queda en texto
inv$Q <- as.numeric(inv$Q)
inv$CREDITO <- as.factor(inv$CREDITO)
inv$UNIDAD <- as.factor(inv$UNIDAD)
str(inv)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   192 obs. of  8 variables:
## $ COD_VIAJE: chr  "10000001" "10000002" "10000003" "10000004" ...
## $ CLIENTE  : chr  "EL PINCHE OBELISCO / Despacho a cliente" "TAQUERIA EL CHINITO |
||Faltante" "TIENDA LA BENDICION / Despacho a cliente" "TAQUERIA EL CHINITO" ...
## $ UBICACION: Factor w/ 2 levels "76001","76002": 2 2 2 2 1 1 2 1 2 2 ...
## $ CANTIDAD : num  1200 1433 1857 339 1644 ...
## $ PILOTO   : chr  "Fernando Mariano Berrio" "Hector Aragones Frutos" "Pedro Alvare
z Parejo" "Angel Valdez Alegria" ...
## $ Q        : num  300 358.2 464.2 84.8 411 ...
## $ CREDITO  : Factor w/ 3 levels "30","60","90": 1 3 2 1 1 1 3 2 1 3 ...
## $ UNIDAD   : Factor w/ 3 levels "Camion Grande",...: 1 1 1 3 1 1 1 1 1 1 ...
```

Seleccionando y nombrando columnas importantes

```
inv <- inv[,c("COD_VIAJE", "CLIENTE", "UBICACION", "CANTIDAD", "PILOTO", "Q", "CREDIT
O", "UNIDAD")]
#names(inv) <- c(vector con nombres de columnas)
```

Funcion substr()

Agregando columna MES y ANIO

```
inv$MES <- as.factor(substr(files[1],1,2))
inv$ANIO <- as.factor(substr(files[1],4,7))
str(inv)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   192 obs. of  10 variables:
## $ COD_VIAJE: chr  "10000001" "10000002" "10000003" "10000004" ...
## $ CLIENTE  : chr  "EL PINCHE OBELISCO / Despacho a cliente" "TAQUERIA EL CHINITO |
||Faltante" "TIENDA LA BENDICION / Despacho a cliente" "TAQUERIA EL CHINITO" ...
## $ UBICACION: Factor w/ 2 levels "76001","76002": 2 2 2 2 1 1 2 1 2 2 ...
## $ CANTIDAD : num  1200 1433 1857 339 1644 ...
## $ PILOTO   : chr  "Fernando Mariano Berrio" "Hector Aragones Frutos" "Pedro Alvare
z Parejo" "Angel Valdez Alegria" ...
## $ Q        : num  300 358.2 464.2 84.8 411 ...
## $ CREDITO  : Factor w/ 3 levels "30","60","90": 1 3 2 1 1 1 3 2 1 3 ...
## $ UNIDAD   : Factor w/ 3 levels "Camion Grande",...: 1 1 1 3 1 1 1 1 1 1 ...
## $ MES      : Factor w/ 1 level "01": 1 1 1 1 1 1 1 1 1 1 ...
## $ ANIO     : Factor w/ 1 level "2017": 1 1 1 1 1 1 1 1 1 1 ...
```

Ciclo para union de todos los archivos

```
i=2
while(i<=length(files)){
  g <- read_excel(files[i]) #sheet = "sheet1"
  g <- g[,c("COD_VIAJE", "CLIENTE", "UBICACION", "CANTIDAD", "PILOTO", "Q", "CREDIT
0", "UNIDAD")]
  g$MES <- substr(files[i],1,2)
  g$ANIO <- substr(files[i],4,7)
  #names(g) <- c(vector con nombres de columnas)
  inv <- rbind(inv,g)
  i=i+1
}
str(inv)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':   2180 obs. of  10 variables:
## $ COD_VIAJE: chr  "10000001" "10000002" "10000003" "10000004" ...
## $ CLIENTE : chr  "EL PINCHE OBELISCO / Despacho a cliente" "TAQUERIA EL CHINITO |
||Faltante" "TIENDA LA BENDICION / Despacho a cliente" "TAQUERIA EL CHINITO" ...
## $ UBICACION: Factor w/ 2 levels "76001","76002": 2 2 2 2 1 1 2 1 2 2 ...
## $ CANTIDAD : num  1200 1433 1857 339 1644 ...
## $ PILOTO : chr  "Fernando Mariano Berrio" "Hector Aragones Frutos" "Pedro Alvare
z Parejo" "Angel Valdez Alegria" ...
## $ Q : num  300 358.2 464.2 84.8 411 ...
## $ CREDITO : Factor w/ 3 levels "30","60","90": 1 3 2 1 1 1 3 2 1 3 ...
## $ UNIDAD : Factor w/ 3 levels "Camion Grande",...: 1 1 1 3 1 1 1 1 1 1 ...
## $ MES : Factor w/ 11 levels "01","02","03",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ ANIO : Factor w/ 1 level "2017": 1 1 1 1 1 1 1 1 1 1 ...
```

```
head(inv)
```

| COD_VI... | CLIENTE | UBICAC... | CANTI... |
|-----------|---|-----------|----------|
| <chr> | <chr> | <fctr> | <dbl> |
| 10000001 | EL PINCHE OBELISCO / Despacho a cliente | 76002 | 1200 |
| 10000002 | TAQUERIA EL CHINITO Faltante | 76002 | 1433 |
| 10000003 | TIENDA LA BENDICION / Despacho a cliente | 76002 | 1857 |
| 10000004 | TAQUERIA EL CHINITO | 76002 | 339 |
| 10000005 | CHICHARRONERIA EL RICO COLESTEROL Faltante | 76001 | 1644 |
| 10000006 | UBIQUO LABS FALTANTE | 76001 | 1827 |

6 rows | 1-4 of 10 columns

<

>

Resumen de costo por Mes

```
inv %>% dplyr::group_by(MES) %>% dplyr::summarise(`Costo Total` = sum(Q))
```

| MES <fctr> | Costo Total <dbl> |
|-------------------|----------------------|
| 01 | 55416.50 |
| 02 | 56226.50 |
| 03 | 48466.75 |
| 04 | 51707.50 |
| 05 | 60075.00 |
| 06 | 52262.00 |
| 07 | 56682.75 |
| 08 | 53262.50 |
| 09 | 51519.25 |
| 10 | 55179.50 |
| 1-10 of 11 rows | |
| Previous 1 2 Next | |

Resumen de costo por Ubicacion

```
inv %>% dplyr::group_by(UBICACION) %>% dplyr::summarise(`Costo Total` = sum(Q))
```

| UBICACION <fctr> | Costo Total <dbl> |
|---------------------|----------------------|
| 76001 | 298277.2 |
| 76002 | 300571.0 |
| 2 rows | |

Resumen de costo por Unidad

```
inv %>% dplyr::group_by(UNIDAD) %>% dplyr::summarise(`Costo Total` = sum(Q))
```

| UNIDAD <fctr> | Costo Total <dbl> |
|------------------|----------------------|
| | |

| UNIDAD <fctr> | Costo Total <dbl> |
|------------------|----------------------|
| Camion Grande | 455466.5 |
| Camion Pequeño | 112815.2 |
| Panel | 30566.5 |
| 3 rows | |

Resumen de costo por Mes, Ubicacion y Unidad

```
inv %>% dplyr::group_by(MES, UBICACION, UNIDAD) %>% dplyr::summarise(`Costo Total` = sum(Q))
```

| MES <fctr> | UBICACION <fctr> | UNIDAD <fctr> | Costo Total <dbl> |
|-----------------|---------------------|------------------|----------------------|
| 01 | 76001 | Camion Grande | 21941.00 |
| 01 | 76001 | Camion Pequeño | 5112.25 |
| 01 | 76001 | Panel | 1263.25 |
| 01 | 76002 | Camion Grande | 23265.50 |
| 01 | 76002 | Camion Pequeño | 2862.75 |
| 01 | 76002 | Panel | 971.75 |
| 02 | 76001 | Camion Grande | 21637.50 |
| 02 | 76001 | Camion Pequeño | 4530.75 |
| 02 | 76001 | Panel | 1307.75 |
| 02 | 76002 | Camion Grande | 21584.00 |
| 1-10 of 66 rows | | Previous | 1 2 3 4 5 6 7 Next |