Transformación de datos

Michael Cifuentes

30 de noviembre de 2020

## Librerias y carga de bases

library(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

library(tidyr)  
library(stringr)  
library(readr)  
data\_fnl <- read\_csv("https://raw.githubusercontent.com/jorgehsaavedra/20201124-test-convocatoria/main/01.%20transformacion\_datos/data\_fnl.csv")

## Parsed with column specification:  
## cols(  
## FechaContable = col\_date(format = ""),  
## FechaTransaccion = col\_date(format = ""),  
## HoraTransaccion = col\_time(format = ""),  
## Empresa = col\_double(),  
## Paradero = col\_double(),  
## StrParadero = col\_character(),  
## Linea = col\_double(),  
## StrLinea = col\_character(),  
## Bus = col\_double(),  
## TipoValidacion = col\_character(),  
## RutaSae = col\_double(),  
## StrRutaSae = col\_logical(),  
## Cenefa = col\_character()  
## )

data\_org <- read\_csv2("https://raw.githubusercontent.com/jorgehsaavedra/20201124-test-convocatoria/main/01.%20transformacion\_datos/data\_org.csv")

## Using ',' as decimal and '.' as grouping mark. Use read\_delim() for more control.

## Parsed with column specification:  
## cols(  
## `Fecha Clearing` = col\_character(),  
## `D<ed>a Trx` = col\_date(format = ""),  
## `Hora Trx` = col\_time(format = ""),  
## Operador = col\_double(),  
## `Ruta Modificada` = col\_double(),  
## `Linea SAE` = col\_double(),  
## Parada = col\_character(),  
## `Tipo Vehiculo` = col\_character(),  
## Vehiculo = col\_double(),  
## `Tipo de Viaje` = col\_character(),  
## Linea = col\_character()  
## )

#1-Tranformar los nombres de las columnas.

names(data\_org)[2]=iconv(names(data\_org)[2],"ISO-8859-1","UTF-8")  
data\_org$Parada=iconv(data\_org$Parada,"ISO-8859-1","UTF-8")

#2-Renombrar los nombres de las columnas.

names(data\_org)=c("FechaContable","FechaTransaccion","HoraTransaccion","Empresa","RutaSae","Linea",  
 "StrParadero","Tipo Vehiculo","Bus","TipoValidacion","StrLinea")

#3-Crear o eliminar columnas de acuerdo con el dataset final.

data\_org=data\_org %>% mutate(Paradero=NA,StrRutaSae=NA,Cenefa=NA)

#4-Convertir las columnas de fecha en formato “%Y-%m-%d”.

data\_org$FechaContable=as.Date(data\_org$FechaContable,format="%Y-%m-%d")  
data\_org$FechaTransaccion=as.Date(data\_org$FechaTransaccion,format="%Y/%m/%d")

#5-De la columna ‘StrParadero’ obtener solo los digitos.

data\_org$Paradero=gsub(")","",str\_extract(data\_org$StrParadero,"[^(]+[$)]"))  
data\_org$StrParadero=gsub("\\).","",str\_extract(data\_org$StrParadero,"\\).\*"))

#6-De la columna ‘StrLinea’ obtener solo el texto.

data\_org$StrLinea=gsub("\\).","",str\_extract(data\_org$StrLinea,"\\).\*"))

#7-Proceso de guardado de resultado.

data\_org=data\_org %>% select("FechaContable","FechaTransaccion","HoraTransaccion","Empresa","Paradero",  
 "StrParadero","Linea","StrLinea","Bus","TipoValidacion",  
 "RutaSae","StrRutaSae","Cenefa")  
  
save(data\_org,file = "data\_fnl\_michael.csv")