02441 – Lead in for Case Campy

Data for campylobacter in Danish broiler flocks (Chickens for meat) are stored at DTU FOOD and comes from three different sources. The purpose of this exercise is to merge these three dataset into one and at the same time remove records with a long range of errors.

The data are in the following three text files: "campy_pre2002.txt", "campy_2002-2005.csv", "campy_2005-.csv".

Columns in first file:

"JNR"	Journal number
"DYRNR"	Animal number - when multiple animals are tested
"MATR"	Type of sample (Material)
"BAKTFUND"	Blank if no campy and subspecies if positive
"DYREART"	Type of animal
"PRV_DATO"	Date when flock is slaughtered (And tested)
"CHR_NR"	CHR nr (ID) of farm (Recoded)
"AKTVNR"	Account No.
"SEKTION"	Which section in DTU FOOD ordered the sample
"EPINR"	ID of house on farm
"SYSBEM"	Note (ignore)
"ANALYSE"	Which type of analysis was performed (ignore)
"region"	R1 to R8 for different areas in Jutland and Funen

The second and third files have almost the same columns and only new ones will be described.

"Materialeart"	See first file (MATR)
"Jnr"	See first file
"Dyrnr"	(ignore)
"Chrnr"	See first file
"Epi.nr","Epinr"	See first file
"Prvdato","Provedato"	See first file
"Resultat"	Result of test ("POS" or "NEG")
"region"	See first file
"Modtag.Dato"	Date sample received at laboratory (ignore)
"Provenr"	(ignore)
"Tolkning"	Result of test ("POSITIV" or "NEGATIV")

The following eight columns are of interest in a combined file:
"chrnr" "epinr" "jnr" "dyrnr" "matr" "resultat" "prvdato" "region"

Steps in merging and cleaning the files:

1) pre2002: Remove those with SEKTION=="res"

2) pre2002: Only keep those with AKTVNR==5133

3) All files: Valid CHR numbers are 10000 and above

4) Convert dates to common format.
Hint: use "as.Date"

5) get same order of columns to keep and then rename.

6) Some tests are recorded in two files with different JNR!?! (Due to transitions between databases ...) (This step can be skipped at first and handled if time permits)

7) Merge the data using "rbind"

8) Remove records with chrnr<=10000 and those with NA as epinr.
  Hint: Use "!is.na(epinr)"

9) Reduce the levels of resultat to only "POS" or "NEG"

10) Remove records with duplicated jnr (Keep first record)
   Hint: Use "duplicated"

11) Only keep records with "matr" in
   c("Kloaksvaber","Svaberprøve","766","772")

12) Add week number since week one 1998 for each record

13) Only keep those with positive week number

14) Some chrnr should be removed due to various reasons.
Skip!

15) It may be decided only to include data from farms that have delivered more than 10 flocks, as those with less may have a bias. This could also be included in the analysis ...

16) Use "split" to split the data by week

17) Summarize number of flocks slaughtered and number of positive flocks per week.

18) Save your data file!

The End!

You should attach your script as an appendix when you hand in Case Campy.