

Aboriginal tables

14 marks

Tables are an important way to display symbolic numbers. In the document **Tables.pdf** (and perhaps more importantly its source file **Tables.Rmd**) you will find some examples of manipulating tables using the **knitr** R package. Please consult those files (downloading them and opening them from RStudio) and familiarize yourself with the material found there. It will be very helpful to you in undertaking the analysis in this questions.

Here you are going to work on some Statistics Canada data to produce an interesting table. The data are on aboriginal populations taken from a Statistics Canada website <http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo60a-eng.htm>; see <https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-011-x/99-011-x2011001-eng.cfm> for archived information on aboriginal peoples in Canada.

This data is available in a “Comma Separated Values” or “csv” file named “aboriginal.csv”. This should appear in the same location as this file of questions. Download both the “.Rmd” file of questions and the “csv” file of data and place them in the same directory on your machine.

In RStudio check what the current working directory is using the `getwd()` command in R. For example in my case the current working directory is

```
getwd()
```

```
path_concat <- function(path1, pathOrFilename2, dirsep = "/") {
  ## Both path1 and pathOrFilename2 are expected to be strings.
  ## path1 is the directory path preceding pathOrFilename2
  ## and does NOT end with a directory separator like / or \ or :
  ##
  ## Note that the directory separator may be different for you.
  ## The default one here is for UNIX or MAC OS,
  ## whereas "\" is used for windows
  paste(path1, pathOrFilename2, sep = dirsep)
}

# With this you could set up a directory string
# identifying where you store your data for this course.
#
# My data directory is
DataPath <- "/Users/clementwu/Downloads"
# This is a unix style pathname (with \ escaping the blank following it).
#
# In Windows, your data directory path would begin with the drive and
# use the windows directory separator "\" e.g. "C:\path\Data"
#
# To be general in what follows, we also use a generic
# directory separator
dirsep <- "/"
# Change yours as necessary.
```

Now that the directory is identified, we can read the csv file into a data frame and show the results.

```
filename <- path_concat(DataPath, "aboriginal.csv", dirsep = dirsep)
data <- read.table(filename, header = TRUE, sep = ",", row.names = "ID")
# You need to load the "knitr" library to access the "kable" function
library(knitr)

# The first few columns look like
kable(data[, 1:4])
```

	CA	NL	PE	NS
Aboriginal.identity.population	3.7539895	4.6852840	1.2890727	2.6769204
North.American.Indian	2.2343213	1.5511076	0.9127827	1.6875394
Metis	1.2476541	1.2924232	0.2868746	0.8504136
Inuit	0.1615824	0.9418509	0.0223539	0.0359876
Non.aboriginal	96.2459944	95.3157148	98.7109273	97.3236333

- a. (10 marks) Reformat this table to make whatever patterns it contains more easily apprehended. Show each step that you choose to follow by displaying the table that results from each step. Say why you chose to make that step by referencing the rules we had for reformatting tables. Write down a summary of whatever patterns you have uncovered.

Step 1: swap rows and columns to place least variation in columns.

```
t1 <- t(data)
kable(t1)
```

	Aboriginal.identity.population	North.American.Indian	Metis	Inuit	Non.aboriginal
CA	3.753989	2.2343213	1.2476541	0.1615824	96.24599
NL	4.685284	1.5511076	1.2924232	0.9418509	95.31571
PE	1.289073	0.9127827	0.2868746	0.0223539	98.71093
NS	2.676920	1.6875394	0.8504136	0.0359876	97.32363
NB	2.452581	1.7209755	0.5933440	0.0257069	97.54672
PQ	1.458128	0.8752801	0.3762824	0.1472585	98.54180
ON	2.015938	1.3167876	0.6119016	0.0169176	97.98410
MB	15.473549	8.8785768	6.3347199	0.0498450	84.52601
SK	14.875504	9.5822194	5.0448184	0.0225402	85.12450
AB	5.784535	2.9872357	2.6254816	0.0494418	94.21546
BC	4.812383	3.1803573	1.4589932	0.0195121	95.18762
YT	25.107651	20.8015899	2.6498841	0.8446505	74.90891
NT	50.255723	30.7842182	8.7189479	10.1315149	49.73210
NU	84.961637	0.3410060	0.4433078	84.0068201	15.02131

Step 2: change numbers to two decimals to reduce number complexity

```
kable(t1, digits=2)
```

	Aboriginal.identity.population	North.American.Indian	Metis	Inuit	Non.aboriginal
CA	3.75	2.23	1.25	0.16	96.25
NL	4.69	1.55	1.29	0.94	95.32
PE	1.29	0.91	0.29	0.02	98.71
NS	2.68	1.69	0.85	0.04	97.32
NB	2.45	1.72	0.59	0.03	97.55
PQ	1.46	0.88	0.38	0.15	98.54
ON	2.02	1.32	0.61	0.02	97.98
MB	15.47	8.88	6.33	0.05	84.53
SK	14.88	9.58	5.04	0.02	85.12
AB	5.78	2.99	2.63	0.05	94.22
BC	4.81	3.18	1.46	0.02	95.19
YT	25.11	20.80	2.65	0.84	74.91
NT	50.26	30.78	8.72	10.13	49.73
NU	84.96	0.34	0.44	84.01	15.02

Step3: change heading to simple labels and add percentage sign. This can make headings more clear to the readers.

```
colnames(t1) <- c('Aboriginal (%)', 'North American Indian (%)', 'Metis (%)', 'Inuit (%)', 'Non aboriginal (%)')
kable(t1, digits=2)
```

	Aboriginal (%)	North American Indian (%)	Metis (%)	Inuit (%)	Non aboriginal (%)
CA	3.75	2.23	1.25	0.16	96.25
NL	4.69	1.55	1.29	0.94	95.32
PE	1.29	0.91	0.29	0.02	98.71
NS	2.68	1.69	0.85	0.04	97.32
NB	2.45	1.72	0.59	0.03	97.55
PQ	1.46	0.88	0.38	0.15	98.54
ON	2.02	1.32	0.61	0.02	97.98
MB	15.47	8.88	6.33	0.05	84.53
SK	14.88	9.58	5.04	0.02	85.12
AB	5.78	2.99	2.63	0.05	94.22
BC	4.81	3.18	1.46	0.02	95.19
YT	25.11	20.80	2.65	0.84	74.91
NT	50.26	30.78	8.72	10.13	49.73
NU	84.96	0.34	0.44	84.01	15.02

Step4: change the alignments in columns to reduce the space between values.

```
kable(t1, digits=2, align='c')
```

	Aboriginal (%)	North American Indian (%)	Metis (%)	Inuit (%)	Non aboriginal (%)
CA	3.75	2.23	1.25	0.16	96.25
NL	4.69	1.55	1.29	0.94	95.32
PE	1.29	0.91	0.29	0.02	98.71
NS	2.68	1.69	0.85	0.04	97.32
NB	2.45	1.72	0.59	0.03	97.55
PQ	1.46	0.88	0.38	0.15	98.54

	Aboriginal (%)	North American Indian (%)	Metis (%)	Inuit (%)	Non aboriginal (%)
ON	2.02	1.32	0.61	0.02	97.98
MB	15.47	8.88	6.33	0.05	84.53
SK	14.88	9.58	5.04	0.02	85.12
AB	5.78	2.99	2.63	0.05	94.22
BC	4.81	3.18	1.46	0.02	95.19
YT	25.11	20.80	2.65	0.84	74.91
NT	50.26	30.78	8.72	10.13	49.73
NU	84.96	0.34	0.44	84.01	15.02

Step5: delete the 'CA' row as it is national statistics but others are provincial.

```
t2 <- t1[-1,]
kable(t2, digits=2, align='c')
```

	Aboriginal (%)	North American Indian (%)	Metis (%)	Inuit (%)	Non aboriginal (%)
NL	4.69	1.55	1.29	0.94	95.32
PE	1.29	0.91	0.29	0.02	98.71
NS	2.68	1.69	0.85	0.04	97.32
NB	2.45	1.72	0.59	0.03	97.55
PQ	1.46	0.88	0.38	0.15	98.54
ON	2.02	1.32	0.61	0.02	97.98
MB	15.47	8.88	6.33	0.05	84.53
SK	14.88	9.58	5.04	0.02	85.12
AB	5.78	2.99	2.63	0.05	94.22
BC	4.81	3.18	1.46	0.02	95.19
YT	25.11	20.80	2.65	0.84	74.91
NT	50.26	30.78	8.72	10.13	49.73
NU	84.96	0.34	0.44	84.01	15.02

Step6: Reorder rows and columns to have rows summaries decrease from top to bottom and column summaries increase from left and right.

```
colorder <- order(colMeans(t2))
roworder <- order(rowMeans(t2), decreasing = TRUE)
t3 <- t2[, colorder][roworder,]
kable(t3, digits=2, align='c')
```

	Metis (%)	North American Indian (%)	Inuit (%)	Aboriginal (%)	Non aboriginal (%)
NU	0.44	0.34	84.01	84.96	15.02
NT	8.72	30.78	10.13	50.26	49.73
YT	2.65	20.80	0.84	25.11	74.91
MB	6.33	8.88	0.05	15.47	84.53
SK	5.04	9.58	0.02	14.88	85.12
AB	2.63	2.99	0.05	5.78	94.22
BC	1.46	3.18	0.02	4.81	95.19
NL	1.29	1.55	0.94	4.69	95.32
NS	0.85	1.69	0.04	2.68	97.32
NB	0.59	1.72	0.03	2.45	97.55

	Metis (%)	North American Indian (%)	Inuit (%)	Aboriginal (%)	Non aboriginal (%)
ON	0.61	1.32	0.02	2.02	97.98
PQ	0.38	0.88	0.15	1.46	98.54
PE	0.29	0.91	0.02	1.29	98.71

Summary: From the table, we can see that most of the provinces have huge groups of non-aboriginal, especially for southern part of Canada, The percentages of Non-aboriginal are over 90%. We can see a gradual increase in the percentage of aboriginal From South to North. The percentages of Metis are pretty low across Canada. North American Indian has a dominant percentage in NT and YT, while most of the people living in MU are Inuit.

- b. (4 marks) Note that the category `Aboriginal.identity.population` includes the “Aboriginal groups (North American Indian, Métis and Inuit), multiple Aboriginal responses and Aboriginal responses not included elsewhere”. Replace the data on `Aboriginal.identity.population` by `Other.aboriginal` that is the difference between `Aboriginal.identity.population` and the North American Indian, Métis and Inuit groups. Again, give the table the best presentation and summarize whatever pattern exists.

```
other <- t3[,4]-t3[,3]-t3[,2]-t3[,1]
t4 <- cbind(t3, other)
```

Step1: change ‘other’ column name and add percentage sign to match the format of other columns

```
colnames(t4)[6] <- 'Other Aboriginal (%)'
kable(t4, digits=2, align='c')
```

	Metis (%)	North American Indian (%)	Inuit (%)	Aboriginal (%)	Non aboriginal (%)	Other Aboriginal (%)
NU	0.44	0.34	84.01	84.96	15.02	0.17
NT	8.72	30.78	10.13	50.26	49.73	0.62
YT	2.65	20.80	0.84	25.11	74.91	0.81
MB	6.33	8.88	0.05	15.47	84.53	0.21
SK	5.04	9.58	0.02	14.88	85.12	0.23
AB	2.63	2.99	0.05	5.78	94.22	0.12
BC	1.46	3.18	0.02	4.81	95.19	0.15
NL	1.29	1.55	0.94	4.69	95.32	0.90
NS	0.85	1.69	0.04	2.68	97.32	0.10
NB	0.59	1.72	0.03	2.45	97.55	0.11
ON	0.61	1.32	0.02	2.02	97.98	0.07
PQ	0.38	0.88	0.15	1.46	98.54	0.06
PE	0.29	0.91	0.02	1.29	98.71	0.07

Step2: remove redundant column ‘Aboriginal’

```
t5 <- t4[, -4]
kable(t5, digits=2, align='c')
```

	Metis (%)	North American Indian (%)	Inuit (%)	Non aboriginal (%)	Other Aboriginal (%)
NU	0.44	0.34	84.01	15.02	0.17
NT	8.72	30.78	10.13	49.73	0.62
YT	2.65	20.80	0.84	74.91	0.81
MB	6.33	8.88	0.05	84.53	0.21
SK	5.04	9.58	0.02	85.12	0.23
AB	2.63	2.99	0.05	94.22	0.12
BC	1.46	3.18	0.02	95.19	0.15
NL	1.29	1.55	0.94	95.32	0.90
NS	0.85	1.69	0.04	97.32	0.10
NB	0.59	1.72	0.03	97.55	0.11
ON	0.61	1.32	0.02	97.98	0.07
PQ	0.38	0.88	0.15	98.54	0.06
PE	0.29	0.91	0.02	98.71	0.07

Step3: order columns and rows by their means

```
colorder <- order(colMeans(t5))
roworder <- order(rowMeans(t5), decreasing = TRUE)
t6 <- t5[, colorder][roworder,]
kable(t6, digits=2, align='c')
```

	Other Aboriginal (%)	Metis (%)	North American Indian (%)	Inuit (%)	Non aboriginal (%)
YT	0.81	2.65	20.80	0.84	74.91
NL	0.90	1.29	1.55	0.94	95.32
NS	0.10	0.85	1.69	0.04	97.32
ON	0.07	0.61	1.32	0.02	97.98
AB	0.12	2.63	2.99	0.05	94.22
SK	0.23	5.04	9.58	0.02	85.12
BC	0.15	1.46	3.18	0.02	95.19
PE	0.07	0.29	0.91	0.02	98.71
PQ	0.06	0.38	0.88	0.15	98.54
MB	0.21	6.33	8.88	0.05	84.53
NB	0.11	0.59	1.72	0.03	97.55
NT	0.62	8.72	30.78	10.13	49.73
NU	0.17	0.44	0.34	84.01	15.02

Summary: From the new graph we can see that, NL has the highest percentage of other aboriginal people, while PQ has the least. Northern part of Canada have high percent of other aboriginal, except for NU, which has only 0.17.