

FigureYa21 TCGA2table

Guangchuang Yu

7/21/2018

```
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.5.1
x = read.csv("easy_input.csv")

x1 = x %>% filter(vital_status == "Alive")
x2 = x %>% filter(vital_status == "Dead" & person_neoplasm_cancer_status == "WITH TUMOR")
x3 = x %>% filter(vital_status == "Dead" & person_neoplasm_cancer_status == "TUMOR FREE")

xx = x %>% filter(vital_status != "" & person_neoplasm_cancer_status != "")

s1 = x1 %>% group_by(gender) %>% summarise(alive=n())
s1

## # A tibble: 2 x 2
##   gender alive
##   <fct> <int>
## 1 FEMALE    84
## 2 MALE    202

s2 = x2 %>% group_by(gender) %>% summarise(tumor=n())
s3 = x3 %>% group_by(gender) %>% summarise(tumor_free=n())

## sex

sex = full_join(s1, s2, by='gender') %>% full_join(s3, by='gender') %>% as.data.frame
sex

##   gender alive tumor tumor_free
## 1 FEMALE    84    25         10
## 2  MALE   202    28         18

rn = sex[,1]
sex = sex[,-1]

sex.p = chisq.test(sex)$p.value
print(sex.p)

## [1] 0.03661159

sex$total = rowSums(sex)
cs = colSums(sex)

sex <- rbind(paste0(sex[1,], " (", sprintf("%1.1f\\%", sex[1,]/cs*100), ")"),
  paste0(sex[2,], " (", sprintf("%1.1f\\%", sex[2,]/cs*100), ")")

rownames(sex) = rn
colnames(sex) = paste0(c("Alive", "Dead with tumor", "Dead tumor free", "Total"),
  "\n(n=", cs, ")")
```

```

print(sex)

##           Alive\n(n=286)  Dead with tumor\n(n=53)  Dead tumor free\n(n=28)
## FEMALE  "84 (29.4\\%)"    "25 (47.2\\%)"          "10 (35.7\\%)"
## MALE    "202 (70.6\\%)"   "28 (52.8\\%)"          "18 (64.3\\%)"
##           Total\n(n=367)
## FEMALE  "119 (32.4\\%)"
## MALE    "248 (67.6\\%)"

## age

age_stats <- function(x) {
  res <- x %>% summarise(age = round(mean(age_at_initial_pathologic_diagnosis,na.rm = T), 1),
    sd=round(sd(age_at_initial_pathologic_diagnosis,na.rm = T), 1),
    median=round(median(age_at_initial_pathologic_diagnosis,na.rm = T), 1),
    min=round(min(age_at_initial_pathologic_diagnosis,na.rm = T), 1),
    max=round(max(age_at_initial_pathologic_diagnosis,na.rm = T), 1)
  )
  c("Mean (SD)" = with(res, paste0(age, " (", sd, ")")),
    "Median [MIN, MAX]" = with(res, paste0(median, " [", min, ", ", max, "]"))#,
  )
}

a1 = age_stats(x1)
a2 = age_stats(x2)
a3 = age_stats(x3)
aa = age_stats(xx)

age = cbind(a1, a2) %>% cbind(a3) %>% cbind(aa)

colnames(age) = colnames(sex)
print(age)

##           Alive\n(n=286)  Dead with tumor\n(n=53)
## Mean (SD)           "58 (13.3)"    "63.1 (13.1)"
## Median [MIN, MAX]   "60 [16,84]"    "64 [23,90]"
##           Dead tumor free\n(n=28)  Total\n(n=367)
## Mean (SD)           "67.1 (11.3)"    "59.5 (13.1)"
## Median [MIN, MAX]   "67.5 [35,83]"    "61 [16,90]"

## Stage

stage_stats <- function(x) {
  x %>% filter(stage_event_pathologic_stage != "") %>%
    group_by(stage_event_pathologic_stage) %>% summarise(stage = n())
}

sg1 = stage_stats(x1)
sg2 = stage_stats(x2)
sg3 = stage_stats(x3)
sgx = stage_stats(xx)
sg = full_join(sg1, sg2, by="stage_event_pathologic_stage") %>%
  full_join(sg3, by="stage_event_pathologic_stage") %>%
  full_join(sgx, by="stage_event_pathologic_stage") %>%
  as.data.frame

```

```
rownames(sg) = sg[,1]
sg = sg[,-1]
colnames(sg) = colnames(sex)

print(sg)
```

```
##           Alive\n(n=286) Dead with tumor\n(n=53) Dead tumor free\n(n=28)
## Stage I           135                19                17
## Stage II          71                 7                 6
## Stage III          1                 1                NA
## Stage IIIA         51                10                 4
## Stage IIIB          6                 2                NA
## Stage IIIC          8                 1                NA
## Stage IV            1                NA                NA
## Stage IVA           1                NA                NA
## Stage IVB          NA                 2                NA
##           Total\n(n=367)
## Stage I           164
## Stage II          78
## Stage III          2
## Stage IIIA         62
## Stage IIIB          8
## Stage IIIC          9
## Stage IV            1
## Stage IVA          NA
## Stage IVB          2
```

```
# chisq test
sgx <- sg[, -4]
sgx <- sgx[!apply(sgx, 1, anyNA),]
sg.p = chisq.test(sgx)$p.value

sgv2 = lapply(1:nrow(sg), function(i) ifelse(is.na(sg[i,]), "",
  paste0(sg[i,], " (", sprintf("%.1f\\%", sg[i,]/cs * 100), ")")))) %>%
  do.call(rbind, .)

rownames(sgv2) = rownames(sg)
colnames(sgv2) = colnames(sg)
print(sgv2)
```

```
##           Alive\n(n=286) Dead with tumor\n(n=53) Dead tumor free\n(n=28)
## Stage I   "135 (47.2\\%)" "19 (35.8\\%)" "17 (60.7\\%)"
## Stage II  "71 (24.8\\%)" "7 (13.2\\%)" "6 (21.4\\%)"
## Stage III "1 (0.3\\%)" "1 (1.9\\%)" ""
## Stage IIIA "51 (17.8\\%)" "10 (18.9\\%)" "4 (14.3\\%)"
## Stage IIIB "6 (2.1\\%)" "2 (3.8\\%)" ""
## Stage IIIC "8 (2.8\\%)" "1 (1.9\\%)" ""
## Stage IV  "1 (0.3\\%)" "" ""
## Stage IVA "1 (0.3\\%)" "" ""
## Stage IVB "" "2 (3.8\\%)" ""
##           Total\n(n=367)
## Stage I   "164 (44.7\\%)"
## Stage II  "78 (21.3\\%)"
## Stage III "2 (0.5\\%)"
```

```
## Stage IIIA "62 (16.9\\%)"
## Stage IIIB "8 (2.2\\%)"
## Stage IIIC "9 (2.5\\%)"
## Stage IV "1 (0.3\\%)"
## Stage IVA ""
## Stage IVB "2 (0.5\\%)"

## combine

res = rbind(sex, age) %>% rbind(sgv2) %>% as.data.frame
print(res)

##           Alive\n(n=286) Dead with tumor\n(n=53)
## FEMALE           84 (29.4\\%)           25 (47.2\\%)
## MALE             202 (70.6\\%)           28 (52.8\\%)
## Mean (SD)         58 (13.3)           63.1 (13.1)
## Median [MIN, MAX]  60 [16,84]          64 [23,90]
## Stage I           135 (47.2\\%)          19 (35.8\\%)
## Stage II           71 (24.8\\%)           7 (13.2\\%)
## Stage III           1 (0.3\\%)            1 (1.9\\%)
## Stage IIIA         51 (17.8\\%)          10 (18.9\\%)
## Stage IIIB          6 (2.1\\%)            2 (3.8\\%)
## Stage IIIC          8 (2.8\\%)            1 (1.9\\%)
## Stage IV            1 (0.3\\%)
## Stage IVA            1 (0.3\\%)
## Stage IVB                                2 (3.8\\%)
##           Dead tumor free\n(n=28) Total\n(n=367)
## FEMALE           10 (35.7\\%)          119 (32.4\\%)
## MALE             18 (64.3\\%)          248 (67.6\\%)
## Mean (SD)         67.1 (11.3)          59.5 (13.1)
## Median [MIN, MAX]  67.5 [35,83]          61 [16,90]
## Stage I           17 (60.7\\%)          164 (44.7\\%)
## Stage II           6 (21.4\\%)           78 (21.3\\%)
## Stage III           2 (0.5\\%)
## Stage IIIA          4 (14.3\\%)          62 (16.9\\%)
## Stage IIIB          8 (2.2\\%)
## Stage IIIC          9 (2.5\\%)
## Stage IV            1 (0.3\\%)
## Stage IVA
## Stage IVB                                2 (0.5\\%)

#install.packages("kableExtra")
require(kableExtra)

if (knitr:::is_html_output()) {
  cn = sub("\\n", "<br>", colnames(res))
} else if (knitr:::is_latex_output()) {
  usepackage_latex('makecell')
  usepackage_latex('booktabs')
  cn = linebreak(colnames(res), align="c")
}

res %>%
  kable(booktabs = T, escape = F, caption = "Example Table",
        col.names = cn) %>%
```

Table 1: Example Table

	Alive (n=286)	Dead with tumor (n=53)	Dead tumor free (n=28)	Total (n=367)
Gender*				
FEMALE	84 (29.4%)	25 (47.2%)	10 (35.7%)	119 (32.4%)
MALE	202 (70.6%)	28 (52.8%)	18 (64.3%)	248 (67.6%)
Age				
Mean (SD)	58 (13.3)	63.1 (13.1)	67.1 (11.3)	59.5 (13.1)
Median [MIN, MAX]	60 [16,84]	64 [23,90]	67.5 [35,83]	61 [16,90]
Stage				
Stage I	135 (47.2%)	19 (35.8%)	17 (60.7%)	164 (44.7%)
Stage II	71 (24.8%)	7 (13.2%)	6 (21.4%)	78 (21.3%)
Stage III	1 (0.3%)	1 (1.9%)		2 (0.5%)
Stage IIIA	51 (17.8%)	10 (18.9%)	4 (14.3%)	62 (16.9%)
Stage IIIB	6 (2.1%)	2 (3.8%)		8 (2.2%)
Stage IIIC	8 (2.8%)	1 (1.9%)		9 (2.5%)
Stage IV	1 (0.3%)			1 (0.3%)
Stage IVA	1 (0.3%)			
Stage IVB		2 (3.8%)		2 (0.5%)

*. significant

```
kable_styling(c("striped", "scale_down")) %>%
group_rows("Gender*", 1, 2) %>%
group_rows("Age", 3, 4) %>%
group_rows("Stage", 5, 13) %>%
footnote(general = "significant",
         #general = paste("P-value =", sg.p),
         general_title = "*: ",
         footnote_as_chunk = T, title_format = "italic")
```

Table 2: Example Table

	Alive (n=286)	Dead with tumor (n=53)	Dead tumor free (n=28)	Total (n=367)	P Value
Gender					
FEMALE	84 (29.4%)	25 (47.2%)	10 (35.7%)	119 (32.4%)	0.037*
MALE	202 (70.6%)	28 (52.8%)	18 (64.3%)	248 (67.6%)	
Age					
Mean (SD)	58 (13.3)	63.1 (13.1)	67.1 (11.3)	59.5 (13.1)	
Median [MIN, MAX]	60 [16,84]	64 [23,90]	67.5 [35,83]	61 [16,90]	
Stage					
Stage I	135 (47.2%)	19 (35.8%)	17 (60.7%)	164 (44.7%)	0.574
Stage II	71 (24.8%)	7 (13.2%)	6 (21.4%)	78 (21.3%)	
Stage III	1 (0.3%)	1 (1.9%)		2 (0.5%)	
Stage IIIA	51 (17.8%)	10 (18.9%)	4 (14.3%)	62 (16.9%)	
Stage IIIB	6 (2.1%)	2 (3.8%)		8 (2.2%)	
Stage IIIC	8 (2.8%)	1 (1.9%)		9 (2.5%)	
Stage IV	1 (0.3%)			1 (0.3%)	
Stage IVA	1 (0.3%)				
Stage IVB		2 (3.8%)		2 (0.5%)	

* significant

```

res[["P Value"]] = c("", paste0(round(sex.p, 3), footnote_marker_symbol(1)), ## sex
  rep("", 2), ## age
  rep("", nrow(sg)-1), round(sg.p, 3) ## stage
)
cn <- c(cn, "P Value")
res %>%
  kable(booktabs = T, escape = F, caption = "Example Table",
    col.names = cn) %>%
  kable_styling(c("striped", "scale_down")) %>%
  group_rows("Gender", 1, 2) %>%
  group_rows("Age", 3, 4) %>%
  group_rows("Stage", 5, 13) %>%
  footnote(general = "significant",
    #general = paste("P-value =", sg.p),
    general_title = "*",
    footnote_as_chunk = T, title_format = "italic")

```

sessionInfo()

```
## R version 3.5.0 (2018-04-23)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS High Sierra 10.13.5
##
## Matrix products: default
## BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] zh_CN.UTF-8/zh_CN.UTF-8/zh_CN.UTF-8/C/zh_CN.UTF-8/zh_CN.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] kableExtra_0.9.0 bindrcpp_0.2.2  dplyr_0.7.6
##
## loaded via a namespace (and not attached):
## [1] Rcpp_0.12.17      pillar_1.2.3      compiler_3.5.0
## [4] plyr_1.8.4        bindr_0.1.1       tools_3.5.0
## [7] digest_0.6.15     evaluate_0.10.1   tibble_1.4.2
## [10] viridisLite_0.3.0 pkgconfig_2.0.1   rlang_0.2.1
## [13] cli_1.0.0         rstudioapi_0.7    yaml_2.1.19
## [16] stringr_1.3.1     httr_1.3.1        knitr_1.20
## [19] xml2_1.2.0        hms_0.4.2         rprojroot_1.3-2
## [22] tidyselect_0.2.4  glue_1.2.0        R6_2.2.2
## [25] rmarkdown_1.10    purrr_0.2.5       readr_1.1.1
## [28] magrittr_1.5      backports_1.1.2   scales_0.5.0
## [31] htmltools_0.3.6   assertthat_0.2.0  rvest_0.3.2
## [34] colorspace_1.3-2  utf8_1.1.4        stringi_1.2.3
## [37] munsell_0.5.0     crayon_1.3.4
```