

Analyze ham radio callsign suffixes

Andrew Zimolzak

2022-05-11

What, if any, 3-letter ham radio callsign suffixes do not exist anywhere in FCC's database (presumably because they are banned)? Short answer: FUC, PIS, and SOS. And a few other boring ones. *All* other suffixes are seemingly allowed.

Introduction

Ham radio operators can choose a “vanity” callsign. A usual callsign format is “K5ABC”. The “suffix” is the part of the callsign after the number. Often the suffix is 3 letters long. This is enough letters to make some mildly to very offensive letter combinations. So, which 3-letter combos are problematic enough that they never appear? Data can be downloaded from FCC in a file called `1_amat.zip`.

```
library(here)
library(dplyr)
library(data.table)
library(tidyr)
library(knitr) # kable
```

Input data and show size

```
# message("Reading dat (~8 sec...)")
d_1528k = fread(here('EN.dat'), header = FALSE, sep="|", quote="")
d_100k = fread(here('ENshort.dat'), header = FALSE, sep="|", quote="")
```

```
X = d_1528k
dim(X)
```

```
## [1] 1528034      30
```

```
X %>%
  rename(
    tablename = V1, num = V2, callsign = V5, code = V6, lnum = V7, fullname = V8,
    firstname = V9, middlename = V10, lastname = V11, suffix = V12, address = V16,
    city = V17, state = V18, zip = V19, weird_addr_text = V20, misctext = V21,
    zeronum = V22, othernum = V23, othercode = V24) -> renamed
```

String processing of suffixes

```
renamed %>%
  select(callsign) %>%
  separate(
    callsign, sep='[0-9]', into=c('prefix', 'suffix'), extra='merge'
  ) %>%
  mutate(len_suf = nchar(suffix)) %>%
  filter(len_suf >= 3) %>%
  mutate(
    s1 = substr(suffix, 1, 1),
    s2 = substr(suffix, 2, 2),
    s3 = substr(suffix, 3, 3)
  ) -> presuf

dim(presuf)
```

```
## [1] 1361019      6
```

```
dim(presuf)[1] / dim(renamed)[1]
```

```
## [1] 0.8906994
```

```
hams_per_suff <- dim(presuf)[1] / (26 ^ 3)
hams_per_suff
```

```
## [1] 77.43622
```

The above shows: How many hams have a 3-char suffix, what proportion of hams have a 3-char suffix, and how many hams per suffix (mean).

Count unique suffixes

```
presuf %>%
  arrange(suffix) %>%
  group_by(suffix, s1, s2, s3) %>%
  summarise(count = n()) %>%
  ungroup() %>%
  filter(s3 != '0') -> counts
```

```
## `summarise()` has grouped output by 'suffix', 's1', 's2'. You can override using the `.groups`
## argument.
```

```
26 ^ 3 - dim(counts)[1]
```

```
## [1] 90
```

The above shows how many suffixes are *missing*, or not attested, out of the expected $26^3 = 17576$.

Side analysis: most common suffixes

```
counts %>%
  arrange(desc(count), suffix) %>%
```

```
select(suffix, count) %>%
head(n=35) %>%
kable()
```

suffix	count
USA	245
WTF	239
BSA	204
FOX	197
EMA	194
EOC	188
EMT	186
BOB	184
CAT	182
HAM	182
SAR	181
AAA	177
TOM	175
JOE	174
ARC	173
EMS	172
FLY	171
JIM	169
ABC	165
RON	165
RED	161
DAN	159
SKY	158
RAY	156
ART	155
SAM	154
ROB	153
KEN	150
COP	149
DOC	149
TIM	147
ACE	146
JAY	145
LEE	145
LDS	144

Never-seen suffixes

```
counts %>%
  expand(s1, s2, s3) -> expanded

left_join(expanded, counts, by=c('s1', 's2', 's3')) -> counts_all

dim(expanded)

## [1] 17576      3
```

```
dim(counts_all)
```

```
## [1] 17576      5
```

```
26^3
```

```
## [1] 17576
```

Compare dimension of `expanded` and `counts_all` with the expected 26^3 . Expect to see 17,576 show up in all of those. **Funny thing:** I initially saw 18,252 in the `expanded` data frame. Why? That equals $26 \cdot 26 \cdot 27$. Must be extra weird character in `s1` or `s2` or `s3`. Added `filter() %>%` so now probably **fixed**.

```
counts_all %>%  
  filter(is.na(count)) -> forbidden
```

```
dim(forbidden)[1]
```

```
## [1] 90
```

```
26 ^ 3 - dim(counts)[1]
```

```
## [1] 90
```

The above show the counts of never-seen suffixes. Expect the above two to be equal. But again, weird. I first saw 765 versus 89. It so happens that $765 - 89 = 26^3$. Seems like `s3` is sometimes equal to the string “0”.

Final weird thing: Then I was seeing 90 in `forbidden` versus 89 missing from `counts`, even after filtering out the zero versus letter O thing? Filtered it earlier in the workflow instead, and now it matches.

Diagnostic

```
presuf %>% filter(s3 == 0) %>% kable()
```

prefix	suffix	len_suf	s1	s2	s3
WB	MZ0	3	M	Z	0

I even wonder if this is a typo in the database of “zero” replacing “capital letter O, as in Oscar.”

Final tables

Lots of suffixes starting with “Q” are banned. But *only three others* are banned. Here they are.

```
forbidden %>% filter(s1 != 'Q') %>% kable()
```

s1	s2	s3	suffix	count
F	U	C	NA	NA
P	I	S	NA	NA
S	O	S	NA	NA

So, somewhat surprisingly to me, all of the following are *not* banned and exist at least once: ASS, POO, FUK. Here’s a table of sketchy suffixes, with some very common ones thrown in for comparison.

```
counts %>%
  filter(
    suffix == 'ASS' | suffix == 'POO' | suffix == 'KKK' | suffix == 'FUK' |
    suffix == 'TIT' | suffix == 'GOD' | suffix == 'FCC' | suffix == 'XXX' |
    suffix == 'SEX' | suffix == 'PEE' | suffix == 'KOK' | suffix == 'NSA' |
    suffix == 'DIK' | suffix == 'DIC' | suffix == 'CIA' | suffix == 'ZZZ' |
    suffix == 'ZYX' | suffix == 'USA' | suffix == 'WTF'
  ) %>%
  arrange(desc(count)) %>%
  mutate(obs_over_exp = count / hams_per_suff) %>%
  kable()
```

suffix	s1	s2	s3	count	obs_over_exp
USA	U	S	A	245	3.1638941
WTF	W	T	F	239	3.0864110
ZZZ	Z	Z	Z	132	1.7046287
FCC	F	C	C	128	1.6529733
CIA	C	I	A	121	1.5625763
NSA	N	S	A	113	1.4592654
POO	P	O	O	95	1.2268161
PEE	P	E	E	82	1.0589360
ZYX	Z	Y	X	72	0.9297975
KKK	K	K	K	68	0.8781420
GOD	G	O	D	60	0.7748312
XXX	X	X	X	51	0.6586065
ASS	A	S	S	42	0.5423818
SEX	S	E	X	41	0.5294680
KOK	K	O	K	24	0.3099325
TIT	T	I	T	20	0.2582771
DIC	D	I	C	19	0.2453632
FUK	F	U	K	19	0.2453632
DIK	D	I	K	15	0.1937078

“POO” is one of the few mildly rude ones that is actually *more* popular than expected. Yikes at the substring with 68 callsigns containing it.

Boring Q suffixes

```
forbidden %>%
  filter(s1 == 'Q') %>%
  group_by(s1, s2) %>%
  summarise(sum_q = n()) %>%
  kable()
```

`summarise()` has grouped output by 's1'. You can override using the `.groups`
argument.

s1	s2	sum_q
Q	R	26
Q	S	26

s1	s2	sum_q
Q	T	26
Q	U	9

Those numbers mean you can't have *any* suffix starting with "QR," "QS," or "QT." Plus a few starting with "QU."

```
forbidden %>%
  filter(s1 == 'Q' & s2 == 'U') %>%
  kable()
```

s1	s2	s3	suffix	count
Q	U	A	NA	NA
Q	U	B	NA	NA
Q	U	D	NA	NA
Q	U	J	NA	NA
Q	U	K	NA	NA
Q	U	M	NA	NA
Q	U	Q	NA	NA
Q	U	R	NA	NA
Q	U	W	NA	NA

Diagnostic

```
forb_n <- 3 + 26*3 + 9
forb_n
```

```
## [1] 90
```

```
26^3 - forb_n
```

```
## [1] 17486
```

```
dim(counts)
```

```
## [1] 17486      5
```