

# Data Tidying

Gary Holton

1/16/2019

Load packages for this session (suppress warning message)

```
library(dplyr)
library(tidyr)
```

To call a function from a specific package use: `package_name::function_name()`

## Data Cleaning

Read in data files

```
catch <- read.csv('https://knb.ecoinformatics.org/knb/d1/mn/v2/object/df35b.302.1',
                  stringsAsFactors = FALSE)
head(catch)
```

##	Region	Year	Chinook	Sockeye	Coho	Pink	Chum	All	notesRegCode
## 1	SSE	1886	0	5	0	0	0	5	
## 2	SSE	1887	0	155	0	0	0	155	
## 3	SSE	1888	0	224	16	0	0	240	
## 4	SSE	1889	0	182	11	92	0	285	
## 5	SSE	1890	0	251	42	0	0	292	
## 6	SSE	1891	0	274	24	0	0	298	

```
regions <- read.csv('https://knb.ecoinformatics.org/knb/d1/mn/v2/object/df35b.303.1',
                    stringsAsFactors = FALSE)
head(regions)
```

##	code	mgmtArea	areaClass	regionCode
## 1	GSE	Unallocated Southeast Alaska	mgmtArea	1
## 2	NSE	Northern Southeast Alaska	mgmtArea	1
## 3	SSE	Southern Southeast Alaska	mgmtArea	1
## 4	YAK	Yakutat	mgmtArea	1
## 5	PWSgmt	Prince William Sound Management Area	mgmtArea	2
## 6	BER	Bering River Subarea Copper River Subarea	subarea	2

```
##
## 1                                     Included are Southeastern Alaska catches
## 2 Northern Southern Alaska includes Districts 9 through 16 from summer straight northwest to and inc
## 3
## 4
## 5
## 6
```

Begin data cleanup

- remove marginal sum and notes col
- move from wide to long format

```
catch_long <- catch %>%
  select(-All, -notesRegCode) %>%
```

```
gather(key="species", value="catch", -Year, -Region)
# don't gather on year or region
# could also just specify species columns to gather on
head(catch_long)
```

```
##   Region Year species catch
## 1    SSE 1886 Chinook     0
## 2    SSE 1887 Chinook     0
## 3    SSE 1888 Chinook     0
## 4    SSE 1889 Chinook     0
## 5    SSE 1890 Chinook     0
## 6    SSE 1891 Chinook     0
```

```
catch_cleaned <- catch_long %>%
  rename(catch_thousands = catch)
head(catch_cleaned)
```

```
##   Region Year species catch_thousands
## 1    SSE 1886 Chinook                0
## 2    SSE 1887 Chinook                0
## 3    SSE 1888 Chinook                0
## 4    SSE 1889 Chinook                0
## 5    SSE 1890 Chinook                0
## 6    SSE 1891 Chinook                0
```

## Split-Apply-Combine

### Joins