

Tidying Data

Dr. Luis Malpica

February, 2020

Contents

Introduction	1
Install required packages	1
Load & Tidy data	2
Simple Data Manipulation	8

Introduction

You will look at one table with messy data your task is to make it tidy in R

OK, now let's do some basic manipulations:

With this initial data set you will end up with data on comon names of fish species and counts. The main goal is that you need to work only with data from 2011, also, drop the month column as all these were taken on December so yo do not need that column.

Goals: 1) Create a GitHub repository and link it to a new R Project 2) Create the necessary subfolders and description files 3) Create a new Rmarkdown file (annotate & comment as necessary) 4) Import, explore and if necessary clean up the main data set provided (i.e. import to R, verify types of data, data structure, rename variables, etc.): "CompDec2011_2015.csv" hint: remember 'read.csv' >>> Challenge: This file has a couple of bugs, hint: you can fix them on excel or explore the 'na.strings' option of 'read.csv'

- 5) Eliminate all the rows that have no values (hint: use the na.strings option of read.csv)
- 6) Transform from wide to long format
- 7) One of your new columns will have two values, the first corresponds to a site #, the second to a year, split this column and name appropriately (hint, use: names_sep = "." within your pivot function)
- 8) You need to select data from 2011 only (hint: remember 'filter'?)
- 9) You need to select all but the month column (hint: remember 'select'?) >>> Challenge: pipe it out!

Install required packages

```
# If you don't yet have tidyverse run:  
# install.packages("tidyverse")  
library(tidyverse)
```

Load & Tidy data

```
#Clean up as s start of R session
rm(list = ls())

#Load data
MessyData <- read.csv("data/CompDec2011_2015.csv")

#Explore your data
str(MessyData)

## 'data.frame':    148 obs. of  96 variables:
## $ DISTINCT.SPP.LIST: Factor w/ 147 levels "almaco jack",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Month             : Factor w/ 1 level "December": 1 1 1 1 1 1 1 1 1 1 ...
## $ X119.2011         : Factor w/ 6 levels "", "1", "6", "62",...: 1 1 1 1 1 6 1 1 1 1 ...
## $ E3.2011           : Factor w/ 9 levels "", "1", "19", "2",...: 1 1 9 1 5 1 1 1 1 1 ...
## $ X101.2011         : int  NA NA NA NA 4 NA NA NA NA 3 ...
## $ X100.2011         : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X108.2011         : int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ X106.2011         : int  NA NA NA NA NA NA NA NA NA NA 4 ...
## $ X94.2011          : int  NA NA NA NA NA NA NA NA NA NA 3 ...
## $ X102.2011         : int  NA NA NA NA 20 NA NA NA NA NA ...
## $ X93.2011          : int  NA NA NA NA 31 NA NA NA NA NA ...
## $ X84.2011          : int  1 NA NA NA NA NA NA NA NA NA 3 ...
## $ X55.2011          : int  NA NA NA NA 11 NA NA NA NA 3 ...
## $ X104.2011         : int  2 NA NA NA 2 NA NA NA NA 1 ...
## $ X77.2011          : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X89.2011          : int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ X78.2011          : int  NA NA NA NA NA NA NA NA NA NA 2 ...
## $ X75.2011          : int  NA NA NA NA NA NA NA NA NA NA 3 ...
## $ X76.2011          : int  NA NA NA NA 16 NA NA NA NA 3 ...
## $ X79.2011          : int  NA NA NA NA NA NA NA NA NA NA 2 ...
## $ X91.2011          : int  NA NA NA NA NA NA NA NA NA NA 2 ...
## $ X73.2011          : int  NA NA NA NA 201 NA NA NA NA 1 ...
## $ X90.2011          : int  NA NA NA NA 2 NA NA NA NA NA ...
## $ X86.2011          : int  NA NA NA NA 16 NA NA NA NA 1 ...
## $ f5.2011           : int  NA NA NA NA NA NA NA NA NA NA ...
## $ f4.2011           : int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ X72.2011          : int  NA NA NA NA 21 NA NA NA NA 1 ...
## $ X99.2011          : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X71.2011          : int  NA NA NA NA 64 NA NA NA NA NA ...
## $ X70.2011          : int  NA NA NA NA 42 NA NA NA NA 1 ...
## $ new2.2011         : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X107.2011         : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X74.2011          : int  NA NA NA NA NA NA NA NA NA NA 4 ...
## $ X112.2011         : int  NA NA NA NA 34 NA NA NA NA 1 ...
## $ X106.2012         : int  NA NA NA NA 10 NA NA NA NA NA ...
## $ X94.2012          : int  1 NA NA NA 1 NA NA NA NA 1 ...
## $ X104.2012         : int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ X70.2012          : int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ X71.2012          : int  NA NA NA NA 27 NA NA NA NA 6 ...
## $ X119.2012         : int  10 NA NA NA NA NA NA NA NA 21 ...
## $ E3.2012           : int  1 NA NA NA 4 NA NA NA NA 3 ...
```

```

## $ f4.2012      : int  4 NA NA NA NA NA NA NA NA NA 8 ...
## $ f5.2012      : int  NA NA NA NA NA 11 NA NA NA NA 4 ...
## $ X107.2012     : int  NA NA NA NA NA NA NA NA NA NA 5 ...
## $ New2.2012     : int  NA NA NA NA NA 10 NA NA NA NA 2 ...
## $ X99.2012      : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X86.2012      : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X89.2012      : int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ X99.2013      : int  NA NA NA NA NA NA NA NA NA NA 2 ...
## $ X89.2013      : int  NA NA NA NA NA 21 NA NA NA NA 1 ...
## $ F4.2013       : int  NA 1 NA NA NA NA NA NA NA 1 ...
## $ New2.2013     : int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ F5.2013       : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X119.2013     : int  1 2 NA NA NA NA NA NA NA 2 ...
## $ E3.2013       : int  NA 1 NA NA NA NA NA NA NA 1 ...
## $ X108.2013     : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X102.2013     : int  2 NA NA NA NA NA NA NA NA NA ...
## $ X107.2013     : int  NA NA NA NA NA 1 NA NA NA NA ...
## $ X90.2013      : int  NA 8 NA NA 5 NA NA NA NA NA ...
## $ X84.2013      : int  NA NA NA NA NA NA NA NA NA NA 4 ...
## $ X94.2013      : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X78.2013      : int  NA NA NA NA NA 25 NA NA NA NA ...
## $ X73.2013      : int  NA 5 NA NA 3 NA NA NA NA NA ...
## $ X71.2013      : int  NA NA NA NA NA 14 NA NA NA NA ...
## $ X71.2014      : int  NA NA NA NA NA 12 NA NA NA NA 1 ...
## $ X73.2014      : int  NA 1 NA NA 100 NA NA NA NA NA ...
## $ X78.2014      : int  NA NA NA NA NA 20 NA NA NA NA ...
## $ X84.2014      : int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ X89.2014      : int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ X90.2014      : int  NA NA NA NA NA 60 NA NA NA NA ...
## $ X94.2014      : int  NA NA NA NA NA NA NA NA NA NA 2 ...
## $ X99.2014      : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X102.2014     : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X107.2014     : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X108.2014     : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X119.2014     : int  NA NA NA NA NA 10 NA NA NA NA ...
## $ E3.2014       : int  NA NA NA NA NA NA NA NA NA NA ...
## $ F4.2014       : int  NA NA NA NA NA NA NA NA NA NA ...
## $ F5.2014       : int  6 NA NA NA NA NA NA NA NA 1 ...
## $ New2.2014     : int  NA NA NA NA NA 8 NA NA NA NA ...
## $ X108.2015     : int  NA 5 NA NA 2 NA NA NA NA 3 ...
## $ X73.2015      : int  NA NA NA NA NA 30 NA NA NA NA 2 ...
## $ X78.2015      : int  NA NA NA NA NA 6 NA NA NA NA ...
## $ X89.2015      : int  NA NA NA NA NA NA NA NA NA NA 2 ...
## $ X90.2015      : int  NA NA NA NA NA 21 NA NA NA NA ...
## $ X94.2015      : int  1 NA NA NA NA NA NA NA NA NA ...
## $ X102.2015     : int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ X71.2015      : int  NA 3 NA NA NA NA NA NA NA ...
## $ X107.2015     : int  NA NA NA NA NA NA NA NA NA 2 ...
## $ E3.2015       : int  NA NA NA NA NA 11 NA NA NA NA 1 ...
## $ X119.2015     : int  NA NA NA NA NA 18 NA NA NA NA ...
## $ X99.2015      : int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ F5.2015       : int  NA NA NA NA NA NA NA NA NA 2 ...
## $ F4.2015       : int  5 NA NA NA 12 NA NA NA NA 1 ...
## $ new2.2015     : int  NA NA NA NA NA NA NA NA NA NA ...

```

```
## $ X84.2015 : int NA NA NA NA 39 NA NA NA NA 3 ...
```

```
#I would change the variable name of first column
```

```
MessyData <- rename(MessyData, CommName = DISTINCT.SPP.LIST)
str(MessyData)
```

```
## 'data.frame': 148 obs. of 96 variables:
## $ CommName : Factor w/ 147 levels "almaco jack",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Month : Factor w/ 1 level "December": 1 1 1 1 1 1 1 1 1 1 ...
## $ X119.2011: Factor w/ 6 levels "", "1", "6", "62",...: 1 1 1 1 1 6 1 1 1 1 ...
## $ E3.2011 : Factor w/ 9 levels "", "1", "19", "2",...: 1 1 9 1 5 1 1 1 1 1 ...
## $ X101.2011: int NA NA NA NA 4 NA NA NA NA 3 ...
## $ X100.2011: int NA NA NA NA NA NA NA NA NA NA ...
## $ X108.2011: int NA NA NA NA NA NA NA NA NA 1 ...
## $ X106.2011: int NA NA NA NA NA NA NA NA NA 4 ...
## $ X94.2011 : int NA NA NA NA NA NA NA NA NA 3 ...
## $ X102.2011: int NA NA NA NA 20 NA NA NA NA NA ...
## $ X93.2011 : int NA NA NA NA 31 NA NA NA NA NA ...
## $ X84.2011 : int 1 NA NA NA NA NA NA NA NA 3 ...
## $ X55.2011 : int NA NA NA NA 11 NA NA NA NA 3 ...
## $ X104.2011: int 2 NA NA NA 2 NA NA NA NA 1 ...
## $ X77.2011 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X89.2011 : int NA NA NA NA NA NA NA NA NA 1 ...
## $ X78.2011 : int NA NA NA NA NA NA NA NA NA 2 ...
## $ X75.2011 : int NA NA NA NA NA NA NA NA NA 3 ...
## $ X76.2011 : int NA NA NA NA 16 NA NA NA NA 3 ...
## $ X79.2011 : int NA NA NA NA NA NA NA NA NA 2 ...
## $ X91.2011 : int NA NA NA NA NA NA NA NA NA 2 ...
## $ X73.2011 : int NA NA NA NA 201 NA NA NA NA 1 ...
## $ X90.2011 : int NA NA NA NA 2 NA NA NA NA NA ...
## $ X86.2011 : int NA NA NA NA 16 NA NA NA NA 1 ...
## $ f5.2011 : int NA NA NA NA NA NA NA NA NA NA ...
## $ f4.2011 : int NA NA NA NA NA NA NA NA NA 1 ...
## $ X72.2011 : int NA NA NA NA 21 NA NA NA NA 1 ...
## $ X99.2011 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X71.2011 : int NA NA NA NA 64 NA NA NA NA NA ...
## $ X70.2011 : int NA NA NA NA 42 NA NA NA NA 1 ...
## $ new2.2011: int NA NA NA NA NA NA NA NA NA NA ...
## $ X107.2011: int NA NA NA NA NA NA NA NA NA NA ...
## $ X74.2011 : int NA NA NA NA NA NA NA NA NA 4 ...
## $ X112.2011: int NA NA NA NA 34 NA NA NA NA 1 ...
## $ X106.2012: int NA NA NA NA 10 NA NA NA NA NA ...
## $ X94.2012 : int 1 NA NA NA 1 NA NA NA NA 1 ...
## $ X104.2012: int NA NA NA NA NA NA NA NA NA 1 ...
## $ X70.2012 : int NA NA NA NA NA NA NA NA NA 1 ...
## $ X71.2012 : int NA NA NA NA 27 NA NA NA NA 6 ...
## $ X119.2012: int 10 NA NA NA NA NA NA NA NA 21 ...
## $ E3.2012 : int 1 NA NA NA 4 NA NA NA NA 3 ...
## $ f4.2012 : int 4 NA NA NA NA NA NA NA NA 8 ...
## $ f5.2012 : int NA NA NA NA 11 NA NA NA NA 4 ...
## $ X107.2012: int NA NA NA NA NA NA NA NA NA 5 ...
## $ New2.2012: int NA NA NA NA 10 NA NA NA NA 2 ...
## $ X99.2012 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X86.2012 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X89.2012 : int NA NA NA NA NA NA NA NA NA 1 ...
```

```
## $ X99.2013 : int NA NA NA NA NA NA NA NA NA NA 2 ...
## $ X89.2013 : int NA NA NA NA 21 NA NA NA NA 1 ...
## $ F4.2013 : int NA 1 NA NA NA NA NA NA NA 1 ...
## $ New2.2013: int NA NA NA NA NA NA NA NA NA 1 ...
## $ F5.2013 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X119.2013: int 1 2 NA NA NA NA NA NA NA 2 ...
## $ E3.2013 : int NA 1 NA NA NA NA NA NA NA 1 ...
## $ X108.2013: int NA NA NA NA NA NA NA NA NA NA ...
## $ X102.2013: int 2 NA NA NA NA NA NA NA NA NA ...
## $ X107.2013: int NA NA NA NA 1 NA NA NA NA NA ...
## $ X90.2013 : int NA 8 NA NA 5 NA NA NA NA NA ...
## $ X84.2013 : int NA NA NA NA NA NA NA NA NA 4 ...
## $ X94.2013 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X78.2013 : int NA NA NA NA 25 NA NA NA NA NA ...
## $ X73.2013 : int NA 5 NA NA 3 NA NA NA NA NA ...
## $ X71.2013 : int NA NA NA NA 14 NA NA NA NA NA ...
## $ X71.2014 : int NA NA NA NA 12 NA NA NA NA 1 ...
## $ X73.2014 : int NA 1 NA NA 100 NA NA NA NA NA ...
## $ X78.2014 : int NA NA NA NA 20 NA NA NA NA NA ...
## $ X84.2014 : int NA NA NA NA NA NA NA NA NA 1 ...
## $ X89.2014 : int NA NA NA NA NA NA NA NA NA 1 ...
## $ X90.2014 : int NA NA NA NA 60 NA NA NA NA NA ...
## $ X94.2014 : int NA NA NA NA NA NA NA NA NA 2 ...
## $ X99.2014 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X102.2014: int NA NA NA NA NA NA NA NA NA NA ...
## $ X107.2014: int NA NA NA NA NA NA NA NA NA NA ...
## $ X108.2014: int NA NA NA NA NA NA NA NA NA NA ...
## $ X119.2014: int NA NA NA NA 10 NA NA NA NA NA ...
## $ E3.2014 : int NA NA NA NA NA NA NA NA NA NA ...
## $ F4.2014 : int NA NA NA NA NA NA NA NA NA NA ...
## $ F5.2014 : int 6 NA NA NA NA NA NA NA NA 1 ...
## $ New2.2014: int NA NA NA NA 8 NA NA NA NA NA ...
## $ X108.2015: int NA 5 NA NA 2 NA NA NA NA 3 ...
## $ X73.2015 : int NA NA NA NA 30 NA NA NA NA 2 ...
## $ X78.2015 : int NA NA NA NA 6 NA NA NA NA NA ...
## $ X89.2015 : int NA NA NA NA NA NA NA NA NA 2 ...
## $ X90.2015 : int NA NA NA NA 21 NA NA NA NA NA ...
## $ X94.2015 : int 1 NA NA NA NA NA NA NA NA NA ...
## $ X102.2015: int NA NA NA NA NA NA NA NA NA 1 ...
## $ X71.2015 : int NA 3 NA NA NA NA NA NA NA NA ...
## $ X107.2015: int NA NA NA NA NA NA NA NA NA 2 ...
## $ E3.2015 : int NA NA NA NA 11 NA NA NA NA 1 ...
## $ X119.2015: int NA NA NA NA 18 NA NA NA NA NA ...
## $ X99.2015 : int NA NA NA NA NA NA NA NA NA 1 ...
## $ F5.2015 : int NA NA NA NA NA NA NA NA NA 2 ...
## $ F4.2015 : int 5 NA NA NA 12 NA NA NA NA 1 ...
## $ new2.2015: int NA NA NA NA NA NA NA NA NA NA ...
## $ X84.2015 : int NA NA NA NA 39 NA NA NA NA 3 ...
```

#Also, something looks odd on two columns
`is(MessyData$X119.2011)`

```
## [1] "factor"           "integer"           "oldClass"
## [4] "double"           "numeric"           "vector"
## [7] "data.frameRowLabels"
```

```

is(MessyData$E3.2011)

## [1] "factor"          "integer"          "oldClass"
## [4] "double"           "numeric"          "vector"
## [7] "data.frameRowLabels"

is.factor(MessyData$X119.2011)

## [1] TRUE

is.factor(MessyData$E3.2011)

## [1] TRUE

#There are some character values on these two columns

#You can choose to go ahead and try to go to longer format regardless and see what happens:
#1st Attempt to go from Wide to long format
#TidyData <- pivot_longer(MessyData, X119.2011:X84.2015, names_to = "Site.Year", values_to = "Counts")
#Error points to the different types of data were have so better to remove those character values:

#It may be best to handle this issue on import using the na.strings option of read.csv
#This will replace all "missing" entries wit NA
MessyData <- read.csv("data/CompDec2011_2015.csv", na.strings = "missing")
MessyData <- rename(MessyData, CommName = DISTINCT.SPP.LIST)
str(MessyData)

## 'data.frame':   148 obs. of  96 variables:
## $ CommName : Factor w/ 147 levels "almaco jack",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ Month : Factor w/ 1 level "December": 1 1 1 1 1 1 1 1 1 1 ...
## $ X119.2011: int  NA NA NA NA NA NA NA NA NA NA NA ...
## $ E3.2011 : int  NA NA NA NA 23 NA NA NA NA NA ...
## $ X101.2011: int  NA NA NA NA 4 NA NA NA NA 3 ...
## $ X100.2011: int  NA NA NA NA NA NA NA NA NA NA NA ...
## $ X108.2011: int  NA NA NA NA NA NA NA NA NA NA 1 ...
## $ X106.2011: int  NA NA NA NA NA NA NA NA NA NA 4 ...
## $ X94.2011 : int  NA NA NA NA NA NA NA NA NA NA 3 ...
## $ X102.2011: int  NA NA NA NA 20 NA NA NA NA NA ...
## $ X93.2011 : int  NA NA NA NA 31 NA NA NA NA NA ...
## $ X84.2011 : int  1 NA NA NA NA NA NA NA NA 3 ...
## $ X55.2011 : int  NA NA NA NA 11 NA NA NA NA 3 ...
## $ X104.2011: int  2 NA NA NA 2 NA NA NA NA 1 ...
## $ X77.2011 : int  NA NA NA NA NA NA NA NA NA NA ...
## $ X89.2011 : int  NA NA NA NA NA NA NA NA NA 1 ...
## $ X78.2011 : int  NA NA NA NA NA NA NA NA NA 2 ...
## $ X75.2011 : int  NA NA NA NA NA NA NA NA NA 3 ...
## $ X76.2011 : int  NA NA NA NA 16 NA NA NA NA 3 ...
## $ X79.2011 : int  NA NA NA NA NA NA NA NA NA 2 ...
## $ X91.2011 : int  NA NA NA NA NA NA NA NA NA 2 ...
## $ X73.2011 : int  NA NA NA NA 201 NA NA NA NA 1 ...
## $ X90.2011 : int  NA NA NA NA 2 NA NA NA NA NA ...
## $ X86.2011 : int  NA NA NA NA 16 NA NA NA NA 1 ...
## $ f5.2011 : int  NA NA NA NA NA NA NA NA NA NA ...
## $ f4.2011 : int  NA NA NA NA NA NA NA NA NA 1 ...
## $ X72.2011 : int  NA NA NA NA 21 NA NA NA NA 1 ...
## $ X99.2011 : int  NA NA NA NA NA NA NA NA NA NA ...

```

```

## $ X71.2011 : int NA NA NA NA 64 NA NA NA NA NA ...
## $ X70.2011 : int NA NA NA NA 42 NA NA NA NA 1 ...
## $ new2.2011: int NA NA NA NA NA NA NA NA NA NA ...
## $ X107.2011: int NA NA NA NA NA NA NA NA NA NA ...
## $ X74.2011 : int NA NA NA NA NA NA NA NA NA 4 ...
## $ X112.2011: int NA NA NA NA 34 NA NA NA NA 1 ...
## $ X106.2012: int NA NA NA NA 10 NA NA NA NA NA ...
## $ X94.2012 : int 1 NA NA NA 1 NA NA NA NA 1 ...
## $ X104.2012: int NA NA NA NA NA NA NA NA NA 1 ...
## $ X70.2012 : int NA NA NA NA NA NA NA NA NA 1 ...
## $ X71.2012 : int NA NA NA NA 27 NA NA NA NA 6 ...
## $ X119.2012: int 10 NA NA NA NA NA NA NA NA 21 ...
## $ E3.2012 : int 1 NA NA NA 4 NA NA NA NA 3 ...
## $ f4.2012 : int 4 NA NA NA NA NA NA NA NA 8 ...
## $ f5.2012 : int NA NA NA NA 11 NA NA NA NA 4 ...
## $ X107.2012: int NA NA NA NA NA NA NA NA NA 5 ...
## $ New2.2012: int NA NA NA NA 10 NA NA NA NA 2 ...
## $ X99.2012 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X86.2012 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X89.2012 : int NA NA NA NA NA NA NA NA NA 1 ...
## $ X99.2013 : int NA NA NA NA NA NA NA NA NA 2 ...
## $ X89.2013 : int NA NA NA NA 21 NA NA NA NA 1 ...
## $ F4.2013 : int NA 1 NA NA NA NA NA NA NA 1 ...
## $ New2.2013: int NA NA NA NA NA NA NA NA NA 1 ...
## $ F5.2013 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X119.2013: int 1 2 NA NA NA NA NA NA NA 2 ...
## $ E3.2013 : int NA 1 NA NA NA NA NA NA NA 1 ...
## $ X108.2013: int NA NA NA NA NA NA NA NA NA NA ...
## $ X102.2013: int 2 NA NA NA NA NA NA NA NA NA ...
## $ X107.2013: int NA NA NA NA 1 NA NA NA NA NA ...
## $ X90.2013 : int NA 8 NA NA 5 NA NA NA NA NA ...
## $ X84.2013 : int NA NA NA NA NA NA NA NA NA 4 ...
## $ X94.2013 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X78.2013 : int NA NA NA NA 25 NA NA NA NA NA ...
## $ X73.2013 : int NA 5 NA NA 3 NA NA NA NA NA ...
## $ X71.2013 : int NA NA NA NA 14 NA NA NA NA NA ...
## $ X71.2014 : int NA NA NA NA 12 NA NA NA NA 1 ...
## $ X73.2014 : int NA 1 NA NA 100 NA NA NA NA NA ...
## $ X78.2014 : int NA NA NA NA 20 NA NA NA NA NA ...
## $ X84.2014 : int NA NA NA NA NA NA NA NA NA 1 ...
## $ X89.2014 : int NA NA NA NA NA NA NA NA NA 1 ...
## $ X90.2014 : int NA NA NA NA 60 NA NA NA NA NA ...
## $ X94.2014 : int NA NA NA NA NA NA NA NA NA 2 ...
## $ X99.2014 : int NA NA NA NA NA NA NA NA NA NA ...
## $ X102.2014: int NA NA NA NA NA NA NA NA NA NA ...
## $ X107.2014: int NA NA NA NA NA NA NA NA NA NA ...
## $ X108.2014: int NA NA NA NA NA NA NA NA NA NA ...
## $ X119.2014: int NA NA NA NA 10 NA NA NA NA NA ...
## $ E3.2014 : int NA NA NA NA NA NA NA NA NA NA ...
## $ F4.2014 : int NA NA NA NA NA NA NA NA NA NA ...
## $ F5.2014 : int 6 NA NA NA NA NA NA NA NA 1 ...
## $ New2.2014: int NA NA NA NA 8 NA NA NA NA NA ...
## $ X108.2015: int NA 5 NA NA 2 NA NA NA NA 3 ...
## $ X73.2015 : int NA NA NA NA 30 NA NA NA NA 2 ...

```

```
## $ X78.2015 : int NA NA NA NA 6 NA NA NA NA NA ...
## $ X89.2015 : int NA NA NA NA NA NA NA NA NA NA 2 ...
## $ X90.2015 : int NA NA NA NA 21 NA NA NA NA NA ...
## $ X94.2015 : int 1 NA NA NA NA NA NA NA NA NA NA ...
## $ X102.2015: int NA NA NA NA NA NA NA NA NA NA 1 ...
## $ X71.2015 : int NA 3 NA NA NA NA NA NA NA NA NA ...
## $ X107.2015: int NA NA NA NA NA NA NA NA NA NA 2 ...
## $ E3.2015 : int NA NA NA NA 11 NA NA NA NA 1 ...
## $ X119.2015: int NA NA NA NA 18 NA NA NA NA NA ...
## $ X99.2015 : int NA NA NA NA NA NA NA NA NA NA 1 ...
## $ F5.2015 : int NA NA NA NA NA NA NA NA NA NA 2 ...
## $ F4.2015 : int 5 NA NA NA 12 NA NA NA NA 1 ...
## $ new2.2015: int NA NA NA NA NA NA NA NA NA NA ...
## $ X84.2015 : int NA NA NA NA 39 NA NA NA NA 3 ...
```

#2nd Attempt to go from Wide to long format

```
TidyData <- pivot_longer(MessyData, X119.2011:X84.2015, names_to = "Site.Year", values_to = "Counts")
str(TidyData)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame': 13912 obs. of 4 variables:
## $ CommName : Factor w/ 147 levels "almaco jack",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Month : Factor w/ 1 level "December": 1 1 1 1 1 1 1 1 1 1 ...
## $ Site.Year: chr "X119.2011" "E3.2011" "X101.2011" "X100.2011" ...
## $ Counts : int NA NA NA NA NA NA NA NA NA NA 1 ...
```

#It worked fine this time, but also need to split in two columns Site and Year
#I would try a 3rd time and split the column from the pivot function:

```
TidyData <- MessyData %>% pivot_longer(cols = X119.2011:X84.2015,
  names_to = c("Site", "Year"),
  names_sep = "\\.",
  values_to = "Counts",
  values_drop_na = TRUE)

str(TidyData) # Check data structure
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame': 771 obs. of 5 variables:
## $ CommName: Factor w/ 147 levels "almaco jack",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ Month : Factor w/ 1 level "December": 1 1 1 1 1 1 1 1 1 1 ...
## $ Site : chr "X84" "X104" "X94" "X119" ...
## $ Year : chr "2011" "2011" "2012" "2012" ...
## $ Counts : int 1 2 1 10 1 4 1 2 6 1 ...
```

#Nice!

Simple Data Manipulation

#OK lets use filter and select to work only with data from 2011 and drop the month column/variable
#Let's pipe it out since there are two functions combined
TidyData <- TidyData %>% filter(Year == 2011) %>% select(-Month)
#Make sure we succeeded
unique(TidyData\$Year)

```
## [1] "2011"
```

```
str(TidyData)
```



```
## Classes 'tbl_df', 'tbl' and 'data.frame':    268 obs. of  4 variables:
## $ CommName: Factor w/ 147 levels "almaco jack",...: 1 1 5 5 5 5 5 5 5 ...
## $ Site     : chr  "X84" "X104" "E3" "X101" ...
## $ Year     : chr  "2011" "2011" "2011" "2011" ...
## $ Counts  : int   1 2 23 4 20 31 11 2 16 201 ...
```

```
#Yesss!!!
```