Lab6

Suixin Jiang

10/20/2019

Exercise 1

```
data(ex0724)
ex0724 %>%
  gather('Canada','Denmark','Netherlands','USA', key = 'Country', value = 'Value') ->
  tidy_ex0724
tidy_ex0724
```

```
##     Year     Country  Value
## 1   1950      Canada     NA
## 2   1951      Canada     NA
## 3   1952      Canada     NA
## 4   1953      Canada     NA
## 5   1954      Canada     NA
## 6   1955      Canada     NA
## 7   1956      Canada     NA
## 8   1957      Canada     NA
## 9   1958      Canada     NA
## 10  1959      Canada     NA
## 11  1960      Canada     NA
## 12  1961      Canada     NA
## 13  1962      Canada     NA
## 14  1963      Canada     NA
## 15  1964      Canada     NA
## 16  1965      Canada     NA
## 17  1966      Canada     NA
## 18  1967      Canada     NA
## 19  1968      Canada     NA
## 20  1969      Canada     NA
## 21  1970      Canada 0.5147
## 22  1971      Canada 0.5153
## 23  1972      Canada 0.5148
## 24  1973      Canada 0.5149
## 25  1974      Canada 0.5141
## 26  1975      Canada 0.5136
## 27  1976      Canada 0.5135
## 28  1977      Canada 0.5145
## 29  1978      Canada 0.5124
## 30  1979      Canada 0.5146
## 31  1980      Canada 0.5136
## 32  1981      Canada 0.5133
## 33  1982      Canada 0.5128
## 34  1983      Canada 0.5145
## 35  1984      Canada 0.5137
## 36  1985      Canada 0.5144
## 37  1986      Canada 0.5123
## 38  1987      Canada 0.5120
```

```
## 39  1988      Canada 0.5122
## 40  1989      Canada 0.5123
## 41  1990      Canada 0.5136
## 42  1991      Canada     NA
## 43  1992      Canada     NA
## 44  1993      Canada     NA
## 45  1994      Canada     NA
## 46  1950     Denmark 0.5120
## 47  1951     Denmark 0.5174
## 48  1952     Denmark 0.5151
## 49  1953     Denmark 0.5175
## 50  1954     Denmark 0.5148
## 51  1955     Denmark 0.5169
## 52  1956     Denmark 0.5153
## 53  1957     Denmark 0.5161
## 54  1958     Denmark 0.5150
## 55  1959     Denmark 0.5139
## 56  1960     Denmark 0.5121
## 57  1961     Denmark 0.5125
## 58  1962     Denmark 0.5122
## 59  1963     Denmark 0.5132
## 60  1964     Denmark 0.5160
## 61  1965     Denmark 0.5148
## 62  1966     Denmark 0.5142
## 63  1967     Denmark 0.5135
## 64  1968     Denmark 0.5164
## 65  1969     Denmark 0.5171
## 66  1970     Denmark 0.5140
## 67  1971     Denmark 0.5170
## 68  1972     Denmark 0.5126
## 69  1973     Denmark 0.5133
## 70  1974     Denmark 0.5127
## 71  1975     Denmark 0.5108
## 72  1976     Denmark 0.5169
## 73  1977     Denmark 0.5144
## 74  1978     Denmark 0.5140
## 75  1979     Denmark 0.5141
## 76  1980     Denmark 0.5125
## 77  1981     Denmark 0.5108
## 78  1982     Denmark 0.5141
## 79  1983     Denmark 0.5117
## 80  1984     Denmark 0.5132
## 81  1985     Denmark 0.5111
## 82  1986     Denmark 0.5142
## 83  1987     Denmark 0.5173
## 84  1988     Denmark 0.5155
## 85  1989     Denmark 0.5132
## 86  1990     Denmark 0.5145
## 87  1991     Denmark 0.5131
## 88  1992     Denmark 0.5143
## 89  1993     Denmark 0.5140
## 90  1994     Denmark 0.5116
## 91  1950 Netherlands 0.5160
## 92  1951 Netherlands 0.5158
```

```
## 93  1952 Netherlands 0.5158
## 94  1953 Netherlands 0.5156
## 95  1954 Netherlands 0.5157
## 96  1955 Netherlands 0.5130
## 97  1956 Netherlands 0.5150
## 98  1957 Netherlands 0.5147
## 99  1958 Netherlands 0.5139
## 100 1959 Netherlands 0.5125
## 101 1960 Netherlands 0.5135
## 102 1961 Netherlands 0.5122
## 103 1962 Netherlands 0.5121
## 104 1963 Netherlands 0.5141
## 105 1964 Netherlands 0.5143
## 106 1965 Netherlands 0.5141
## 107 1966 Netherlands 0.5129
## 108 1967 Netherlands 0.5135
## 109 1968 Netherlands 0.5116
## 110 1969 Netherlands 0.5135
## 111 1970 Netherlands 0.5120
## 112 1971 Netherlands 0.5134
## 113 1972 Netherlands 0.5112
## 114 1973 Netherlands 0.5115
## 115 1974 Netherlands 0.5132
## 116 1975 Netherlands 0.5122
## 117 1976 Netherlands 0.5148
## 118 1977 Netherlands 0.5135
## 119 1978 Netherlands 0.5126
## 120 1979 Netherlands 0.5123
## 121 1980 Netherlands 0.5128
## 122 1981 Netherlands 0.5107
## 123 1982 Netherlands 0.5128
## 124 1983 Netherlands 0.5113
## 125 1984 Netherlands 0.5132
## 126 1985 Netherlands 0.5111
## 127 1986 Netherlands 0.5087
## 128 1987 Netherlands 0.5136
## 129 1988 Netherlands 0.5117
## 130 1989 Netherlands 0.5096
## 131 1990 Netherlands 0.5132
## 132 1991 Netherlands 0.5114
## 133 1992 Netherlands 0.5129
## 134 1993 Netherlands 0.5116
## 135 1994 Netherlands 0.5128
## 136 1950         USA     NA
## 137 1951         USA     NA
## 138 1952         USA     NA
## 139 1953         USA     NA
## 140 1954         USA     NA
## 141 1955         USA     NA
## 142 1956         USA     NA
## 143 1957         USA     NA
## 144 1958         USA     NA
## 145 1959         USA     NA
## 146 1960         USA     NA
```

```
## 147 1961         USA     NA
## 148 1962         USA     NA
## 149 1963         USA     NA
## 150 1964         USA     NA
## 151 1965         USA     NA
## 152 1966         USA     NA
## 153 1967         USA     NA
## 154 1968         USA     NA
## 155 1969         USA     NA
## 156 1970         USA 0.5134
## 157 1971         USA 0.5126
## 158 1972         USA 0.5125
## 159 1973         USA 0.5128
## 160 1974         USA 0.5133
## 161 1975         USA 0.5132
## 162 1976         USA 0.5128
## 163 1977         USA 0.5128
## 164 1978         USA 0.5129
## 165 1979         USA 0.5127
## 166 1980         USA 0.5129
## 167 1981         USA 0.5126
## 168 1982         USA 0.5123
## 169 1983         USA 0.5127
## 170 1984         USA 0.5122
## 171 1985         USA 0.5126
## 172 1986         USA 0.5122
## 173 1987         USA 0.5120
## 174 1988         USA 0.5121
## 175 1989         USA 0.5120
## 176 1990         USA 0.5120
## 177 1991         USA     NA
## 178 1992         USA     NA
## 179 1993         USA     NA
## 180 1994         USA     NA
```

Exercise 2

```
preg <- read_csv(file = './preg.csv')
```

```
## Parsed with column specification:
## cols(
##   name = col_character(),
##   treatmenta = col_double(),
##   treatmentb = col_double()
## )
```

```
preg %>%
  rename(A = treatmenta, B = treatmentb) %>%
  gather('A','B', key = 'Treatment', value = 'Value') ->
  tidy_preg
tidy_preg
```

```
## # A tibble: 6 x 3
##   name         Treatment Value
##   <chr>        <chr>     <dbl>
```

```
## 1 John Smith   A            NA
## 2 Jane Doe     A             4
## 3 Mary Johnson A             6
## 4 John Smith   B            18
## 5 Jane Doe     B             1
## 6 Mary Johnson B             7
```

Exercise 3

```
pew <- read_csv(file = './pew.csv')
```

```
## Parsed with column specification:
## cols(
##   religion = col_character(),
##   `<$10k` = col_double(),
##   `$10-20k` = col_double(),
##   `$20-30k` = col_double(),
##   `$30-40k` = col_double(),
##   `$40-50k` = col_double(),
##   `$50-75k` = col_double(),
##   `$75-100k` = col_double(),
##   `$100-150k` = col_double(),
##   `>150k` = col_double(),
##   `Don't know/refused` = col_double()
## )
```

```
pew %>%
  gather(-religion, key = 'Salary', value = 'Count') ->
  tidy_pew
tidy_pew
```

```
## # A tibble: 180 x 3
##    religion                Salary Count
##    <chr>                   <chr>  <dbl>
##  1 Agnostic                <$10k     27
##  2 Atheist                 <$10k     12
##  3 Buddhist                <$10k     27
##  4 Catholic                <$10k    418
##  5 Don’t know/refused      <$10k     15
##  6 Evangelical Prot        <$10k    575
##  7 Hindu                   <$10k      1
##  8 Historically Black Prot <$10k    228
##  9 Jehovah's Witness       <$10k     20
## 10 Jewish                  <$10k     19
## # ... with 170 more rows
```

Exercise 4

```
tb <- read_csv(file = './tb.csv')
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   iso2 = col_character()
## )
```

```
## See spec(...) for full column specifications.
```

```
tb %>%
  gather(-iso2, -year, key = 'Sex_Age', value = 'Count', na.rm = T) %>%
  separate(Sex_Age, into = c('Sex', 'Age'), sep = 1) ->
  tidy_tb
tidy_tb
```

```
## # A tibble: 35,750 x 5
##    iso2   year Sex   Age   Count
##    <chr> <dbl> <chr> <chr> <dbl>
##  1 AD     2005 m     04        0
##  2 AD     2006 m     04        0
##  3 AD     2008 m     04        0
##  4 AE     2006 m     04        0
##  5 AE     2007 m     04        0
##  6 AE     2008 m     04        0
##  7 AG     2007 m     04        0
##  8 AL     2005 m     04        0
##  9 AL     2006 m     04        1
## 10 AL     2007 m     04        0
## # ... with 35,740 more rows
```

Exercise 5

```
weather <- read_csv(file = './weather.csv')
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   id = col_character(),
##   element = col_character(),
##   d9 = col_logical(),
##   d12 = col_logical(),
##   d18 = col_logical(),
##   d19 = col_logical(),
##   d20 = col_logical(),
##   d21 = col_logical(),
##   d22 = col_logical(),
##   d24 = col_logical()
## )
```

```
## See spec(...) for full column specifications.
```

```
weather %>%
  gather(-id, -year, -month, -element, key = 'day', value = 'temperature', na.rm = T) %>%
  spread(key = 'element', value = 'temperature') %>%
  mutate(day = parse_number(str_replace(day,'d',''))) ->
  tidy_weather
tidy_weather
```

```
## # A tibble: 33 x 6
##    id       year month   day  tmax  tmin
##    <chr>   <dbl> <dbl> <dbl> <dbl> <dbl>
##  1 MX17004  2010     1    30  27.8  14.5
##  2 MX17004  2010     2    11  29.7  13.4
```

```
##  3 MX17004  2010     2     2  27.3  14.4
##  4 MX17004  2010     2    23  29.9  10.7
##  5 MX17004  2010     2     3  24.1  14.4
##  6 MX17004  2010     3    10  34.5  16.8
##  7 MX17004  2010     3    16  31.1  17.6
##  8 MX17004  2010     3     5  32.1  14.2
##  9 MX17004  2010     4    27  36.3  16.7
## 10 MX17004  2010     5    27  33.2  18.2
## # ... with 23 more rows
```

Exercise 6

```
wine <- read_csv2(file = './wine.csv')
```

```
## Using ',' as decimal and '.' as grouping mark. Use read_delim() for more control.
```

```
## Parsed with column specification:
## cols(
##   measure = col_character(),
##   Norway = col_double(),
##   Scotland = col_double(),
##   England = col_double(),
##   Ireland = col_double(),
##   Finland = col_double(),
##   Canada = col_double(),
##   UnitedStates = col_double(),
##   Netherlands = col_double(),
##   NewZealand = col_double(),
##   Denmark = col_double(),
##   Sweden = col_double(),
##   Australia = col_double(),
##   Belgium = col_double(),
##   Germany = col_double(),
##   Austria = col_double(),
##   Switzerland = col_double(),
##   Italy = col_double(),
##   France = col_double()
## )
```

```
wine %>%
  gather(-measure, key = 'country', value = 'value') %>%
  spread(key = 'measure', value = 'value') ->
  tidy_wine
tidy_wine
```

```
## # A tibble: 18 x 3
##    country      mortality  wine
##    <chr>            <dbl> <dbl>
##  1 Australia          9.1   8.3
##  2 Austria            4.7  25.1
##  3 Belgium            5.1  12.6
##  4 Canada             7.8   4.9
##  5 Denmark            5.5   5.9
##  6 England            7.1   3.2
##  7 Finland           10.2   4.3
```

```
##  8 France            2.1  75.9
##  9 Germany           4.7  15.1
## 10 Ireland           6.8   3.4
## 11 Italy             3.2  75.9
## 12 Netherlands       5.9   5.2
## 13 NewZealand        8.9   5.9
## 14 Norway            6.2   2.8
## 15 Scotland          9     3.2
## 16 Sweden            7.1   6.6
## 17 Switzerland       3.1  33.1
## 18 UnitedStates      9.3   5.1
```