Tidy Data

January 31, 2018

1 Tidy Data

```
In [1]: import pandas as pd

        pacientes = ['Ricardo', 'Marielena', 'Miguel']

        a = [ 67, 80, 64 ]
        b = [ 56, 90, 50 ]

        untidy = pd.DataFrame({'pacientes': pacientes, 'a': a, 'b': b })
```

Documentación relacionada: `pandas.melt`
http://pandas.pydata.org/pandas-docs/stable/generated/pandas.melt.html

```
In [2]: def gather(df, key, value, cols):
            id_vars = [col for col in df.columns if col not in cols ]
            id_values = cols
            var_name = key
            value_name = value
            return pd.melt(df, id_vars, id_values, var_name, value_name)
```

```
In [3]: untidy
```

```
Out[3]:     a   b  pacientes
        0  67  56    Ricardo
        1  80  90  Marielena
        2  64  50     Miguel
```

```
In [4]: tidy = gather(untidy, 'medicamento', 'frecuencia_cardiaca', ['a', 'b'])
```

```
In [5]: tidy
```

```
Out[5]:    pacientes medicamento  frecuencia_cardiaca
        0    Ricardo           a                   67
        1  Marielena           a                   80
        2     Miguel           a                   64
        3    Ricardo           b                   56
        4  Marielena           b                   90
        5     Miguel           b                   50
```

1.1 Para saber más:

- https://garrettgman.github.io/tidying/ (R-lang)
- http://www.prometheusresearch.com/good-data-management-practices-for-data-analysis-tidy-data-part-2/ (R-lang)
- https://tomaugspurger.github.io/modern-5-tidy.html (Python+Pandas)
- https://www.ibm.com/developerworks/community/blogs/jfp/entry/Tidy_Data_In_Python?lang=en (Python+Pandas)
- http://connor-johnson.com/2014/08/28/tidyr-and-pandas-gather-and-melt/ (Python+Pandas)
- https://blog.rstudio.org/2014/07/22/introducing-tidyr/ (R-lang)