# Statistics with Julia

**OnlineStats.jl**

- `OnlineStats.jl` provides online algorithms for statistical models.
- Online algorithms are well suited for streaming data or when data is too large to hold in memory.
- Observations are processed one at a time and all algorithms use O(1) memory.

*https://github.com/joshday/OnlineStats.jl*

# Statistics with Julia

```
using OnlineStats
o = Mean()

All OnlineStats can be updated

y = randn(100)

for yi in y
fit!(o, y)
end

# or more simply:
fit!(o, y)
OnlineStats share a common interface

value(o)  # associated value of an OnlineStat
nobs(o)   # number of observations used
```

**What Can OnlineStats Do?** While many estimates can be calculated analytically with an online algorithm, several type rely on stochastic approximation.

```
Summary Statistics

Mean: Mean, Means
Variance: Variance, Variances
Quantiles: QuantileMM, QuantileSGD
Covariance Matrix: CovMatrix
Maximum and Minimum: Extrema
Skewness and Kurtosis: Moments
Sum/Differences: Sum, Sums, Diff, Diffs
Density Estimation
```

```
distributionfit(D, data)
For D in [Beta, Categorical, Cauchy, Gamma, LogNormal, Norm
Gaussian Mixtures: NormalMix
Predictive Modeling
```

Linear Regression: LinReg, StatLearn
Logistic Regression: StatLearn
Poisson Regression: StatLearn
Support Vector Machines: StatLearn
Quantile Regression: StatLearn, QuantRegMM
Huber Loss Regression: StatLearn
L1 Loss Regression: StatLearn

Other
K-Means clustering: KMeans Bootstrapping: BernoulliBootstrap, PoissonBootstrap Approximate count of distinct elements: HyperLogLog