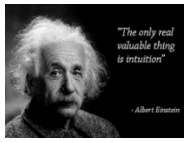
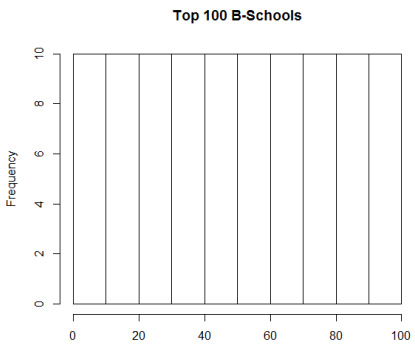
Transformation #8:
Ranks



Rank: Intuition

- Sometimes a variable is important but does not have **enough variance** to capture significant effects in regression models
- Sometimes a variable's **distributions** is **uneven** or non-normal, which is particularly problematic for **small samples**
- A **rank transformation** is a popular **"non-parametric"** statistical approach, which solves some of these problems
- This is done by **sorting** the values and assigning **1** for the smallest value, **2** for the next, etc., or vice versa (highest to smallest)
- The intervals between data points is exactly 1 (i.e., the next value after rank 1 is rank 2), rank transformation has the nice property of producing a **"uniform distribution".** For example the ranks of the top 100 B-Schools distribution looks like this



- **Effect interpretation** → the unit increase or decrease is the rank, not the actual value (e.g., increase rank by 1)
- Some times ranks are **re-scaled** to a **0-1** scale

Tips

`dataName$NewRankVariable.LoHi = rank(originalVariable)` → Creates a new variable ranked from low to high

`dataName$NewRankVariable.HiLo = rank(-originalVariable)` → Using the – sign creates a new variable ranked from high to low

KOGOD SCHOOL
of
BUSINESS