Digital Dark matter

By Shane Greenstein

For CAIDA WIE

The agenda

- GDP is the total flow of final goods and services in an economy over a specific period of time (without double counting).
  - Every developed country measures GDP, capital stocks, labor stock, capital flows and labor flows. Broad effort to understand contributions from various parts.
- **Digital dark matter**: Asset w/o apparent monetary value, that can be replicated without limit, & act as IT input into production.
  - if price is zero, it contributes nothing to GDP
  - Increasing role of "intangible" capital in the economy.
- At this stage: demonstrating **proof of concept.**

One way to motivate it: One of several measurement issues due to a zero price.

	Zero Price in most of the output	Positive price in most of the output
Zero Price in most of the inputs	**Wikipedia**. **Apache/Linux** Foundations **Github**. **And more….**	**Android** phone uses Linux kernel. **Unlicensed spectrum**. - Algorithms for **machine learning**
Positive Price in most of the inputs	**Google/FB.** Free output, ads raise revenue. **Commercial WEB**. See above.	- Manufacturing **iPhone**. **Netflix** pay licenses. Subscribers pay fee. - Carrier paid **license for spectrum**.

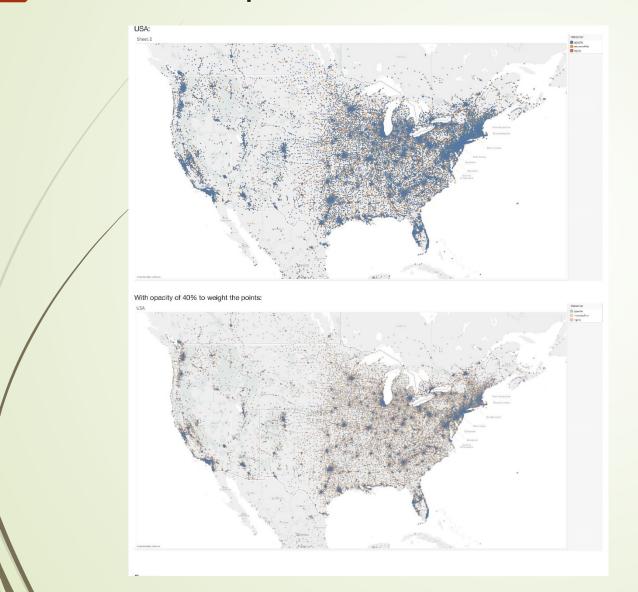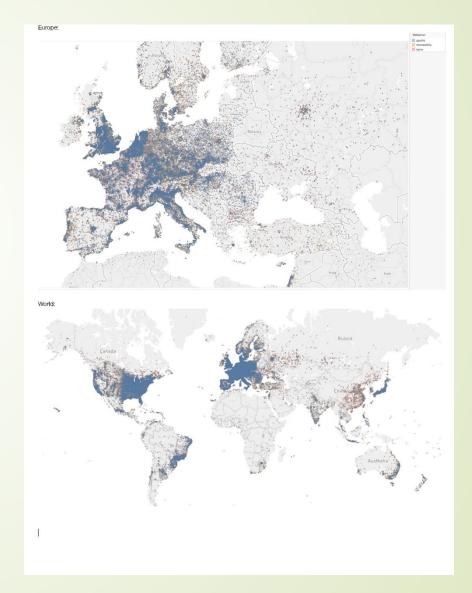# What the first study does (Greenstein and Nagle, 2015)

- Takes a one percent sample of all IP addresses and estimates the total number of Apache servers in the US in 2011. **Proof of concept exercise.**
  - Attach a value to open source servers using standard procedures for "near market goods," as defined by Nordhaus, 2006.
- Apache accounts for a mismeasurement of somewhere b/w $2 billion and $12 billion in software in the US in 2011.
  - **Large.** Equates to b/w 1.3 % & 8.7 % of stock of prepackaged software in private fixed investment in the US
  - **High ROR.** Return on Apache alone would have generated sufficient rate of return to justify investment in Internet R&D by Federal government.
  - **Omission biases.** Large.
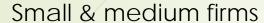 - Attribution bias? No.

What the second study does (Ackermann and Greenstein, 2018)

- Takes a big sample of outward facing servers in the planet in 2012. Apache, nginx, & IIS in 2012. Proof of concept at a **global** level.
  - Locates a server in a country using methods for assigning **long/lat** to IP addresses,. Compare across countries.
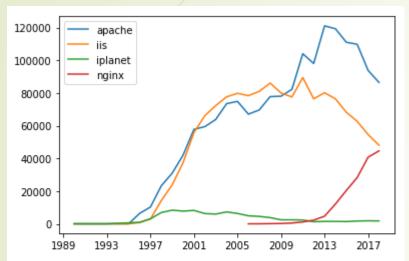
Maps: where servers are located

What the second study does (Ackermann and Greenstein, 2018)

- Takes a big sample of outward facing servers in the planet in 2012. Apache, nginx, & IIS in 2012. Proof of concept at a **global** level.
  - Locates a server in a country using methods for assigning **long/lat** to IP addresses,. Compare across countries.
- Skewed across the world.
  - US has 44% of servers. Next biggest is China, which is six times smaller.
- Analyze who has more and why?
  - Best predictors of server and per-capita servers are the quality of the network and the sophistication of the labor market for technical goods.
  - Rule of law and indications of "failed" institutions do not matter
- Vexing. **Cannot easily forecast open source share**

Third study from Wayback Machine (Greenstein, Murciano-Goroff & Zhuo, 2018)

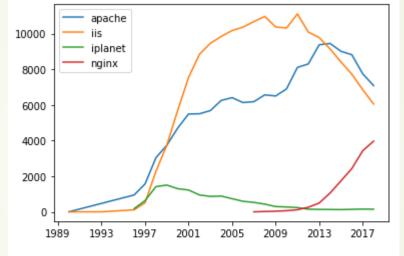| Small & medium firms | Large firms | Largest firms |

Number of observations of unique urls in each year by server type in firms with >=20 & <500 employees
Observations with missing urls or dates are dropped.
Observations before 1990 and after 2018 are dropped.
The observations are all from web scrapes.

Number of observations of unique urls in each year by server type in firms with >=500 & <5000 employees
Observations with missing urls or dates are dropped.
Observations before 1990 and after 2018 are dropped.
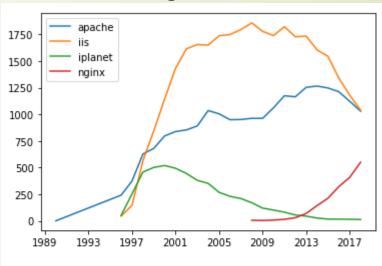The observations are all from web scrapes.

Number of observations of unique urls in each year by server type in firms with >=5000 employees
Observations with missing urls or dates are dropped.
Observations before 1990 and after 2018 are dropped.
The observations are all from web scrapes.

* Preliminary: Usage shares for web sites at 200k+ US firms. A whole asset class for which we have NO investment data. Measured for the first time. Very difficult to clean up.
* Proof of concept. Large (?) mismeasurement of intangible capital inside firms.
* Very different adoption and upgrade behavior across different types of firms.

Open questions:

- Other examples?
  - Frank Nagle has shown that much OS has a direct **productivity** impact on the firms who deploy it.
  - The same mechanisms. **Any open source**? E.g., Linux, Firefox, PERL, PHP, 802.11 software.
  - **Spillovers from university R&D** without licensing. E.g., WWW, browsers, TCP/IP software protocols, etc.
  - **Creative commons:** Wikipedia, Kahn Academy, Github? Unmeasured gains could be large due to measured by audience size.
  - The **spectrum** in Wifi equipment? Standards from IEEE
- Beyond proof of concept? More than a rounding error. Large enough value to matter.