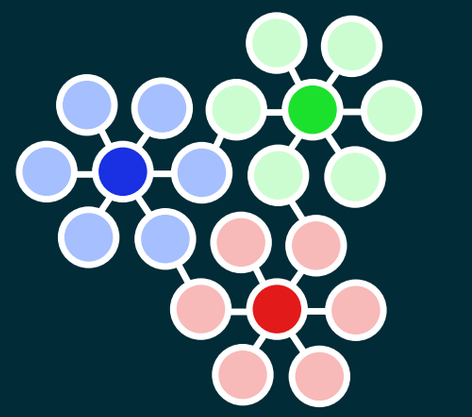
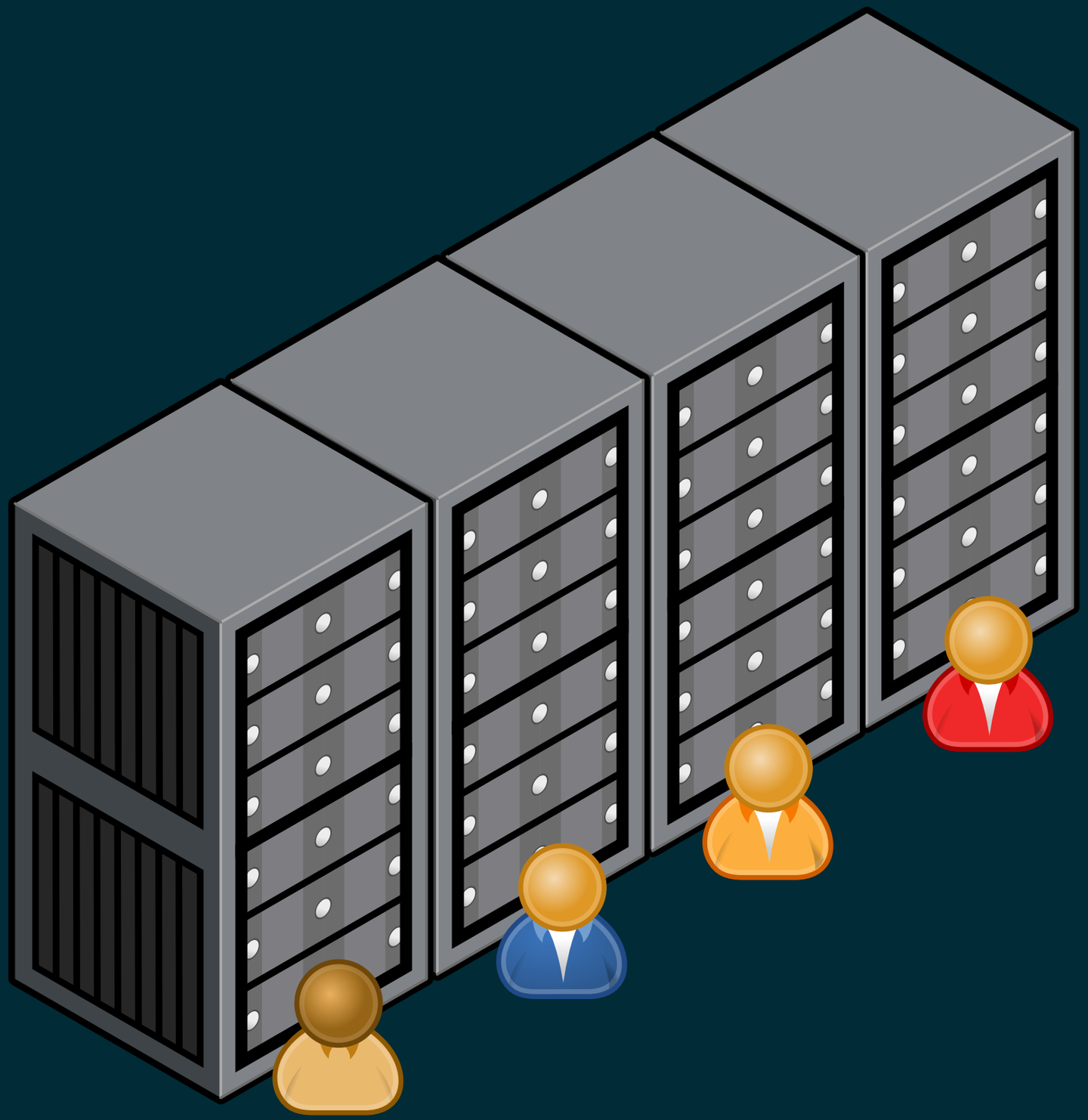
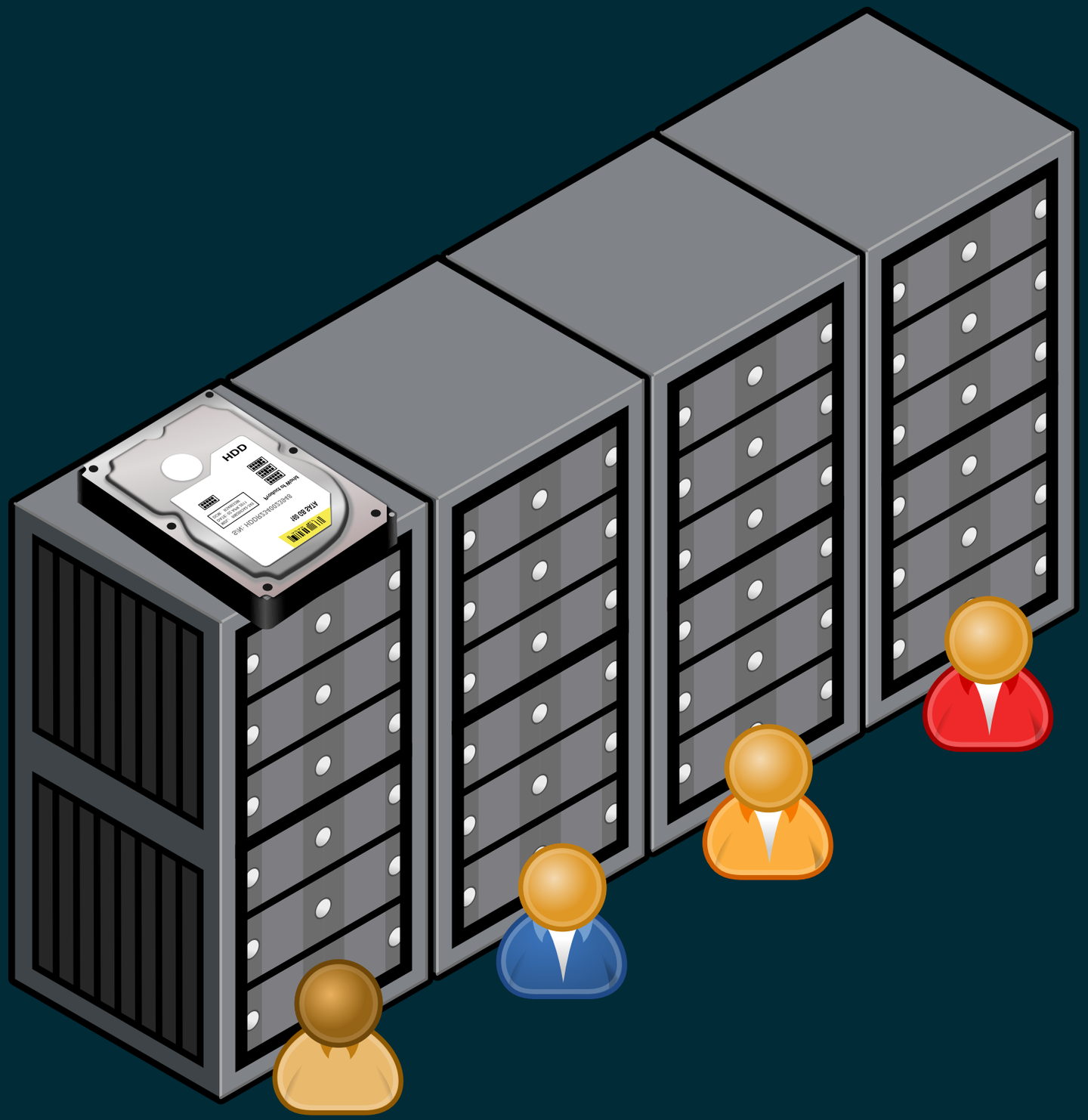
Operational Experiences with Disk Imaging in a Multi-Tenant Datacenter

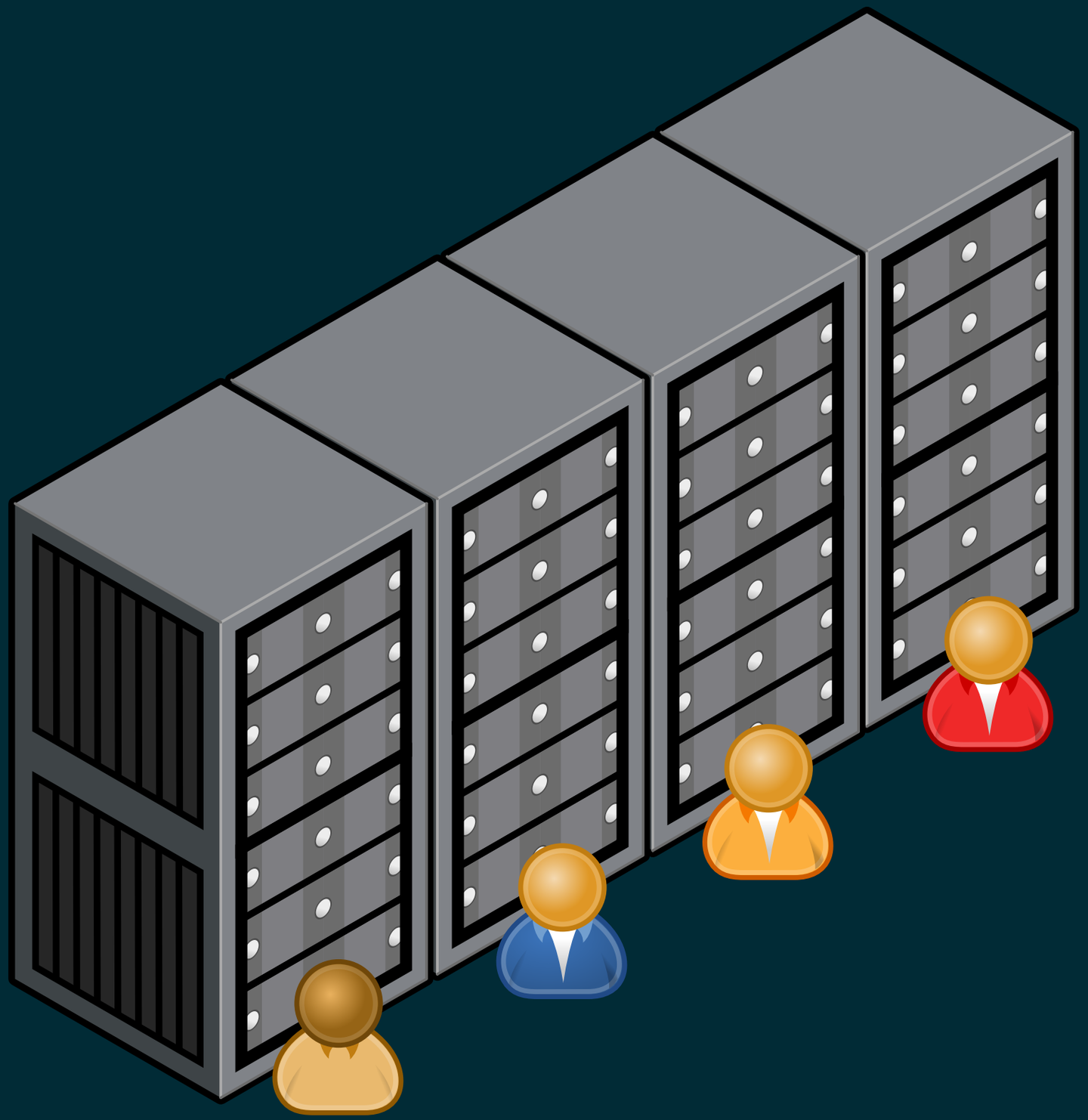
Kevin Atkinson, Gary Wong, and Robert Ricci



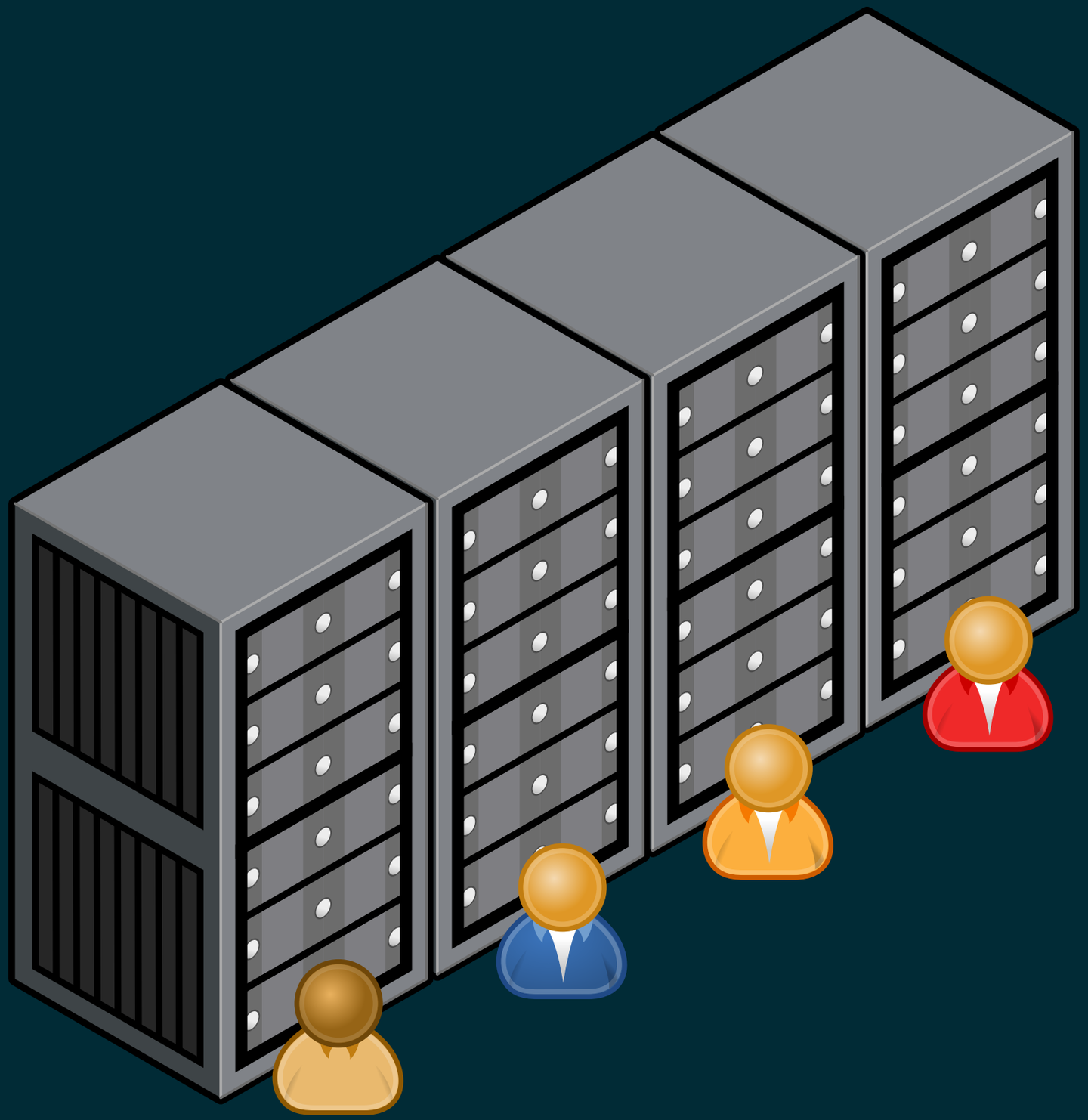
emulab




HDD




HDD




HDD


HDD


HDD


HDD
HDD
HDD


HDD

Properties of disk images and their usage have consequences for:

- Storage
- Caching
- Pre-loading
- Distribution

What does the working set look like?

What does the working set look like?

What do the images themselves look like?

What does the working set look like?

What do the images themselves look like?

What are the key factors in pre-loading?

The dataset

- Four years (2009-2013): 279,972 requests
- Users: 1,301 individuals, 368 organizations
- Unique images: 714
- Emulab
  - ~600 PCs
  - Facility / user image model

User Behavior

“Emulab is a pretty odd beast and its users are even weirder.”

“Emulab is a pretty odd beast and its users are even weirder.”

-Reviewer D

“Emulab is a pretty odd beast and its users are even weirder.”

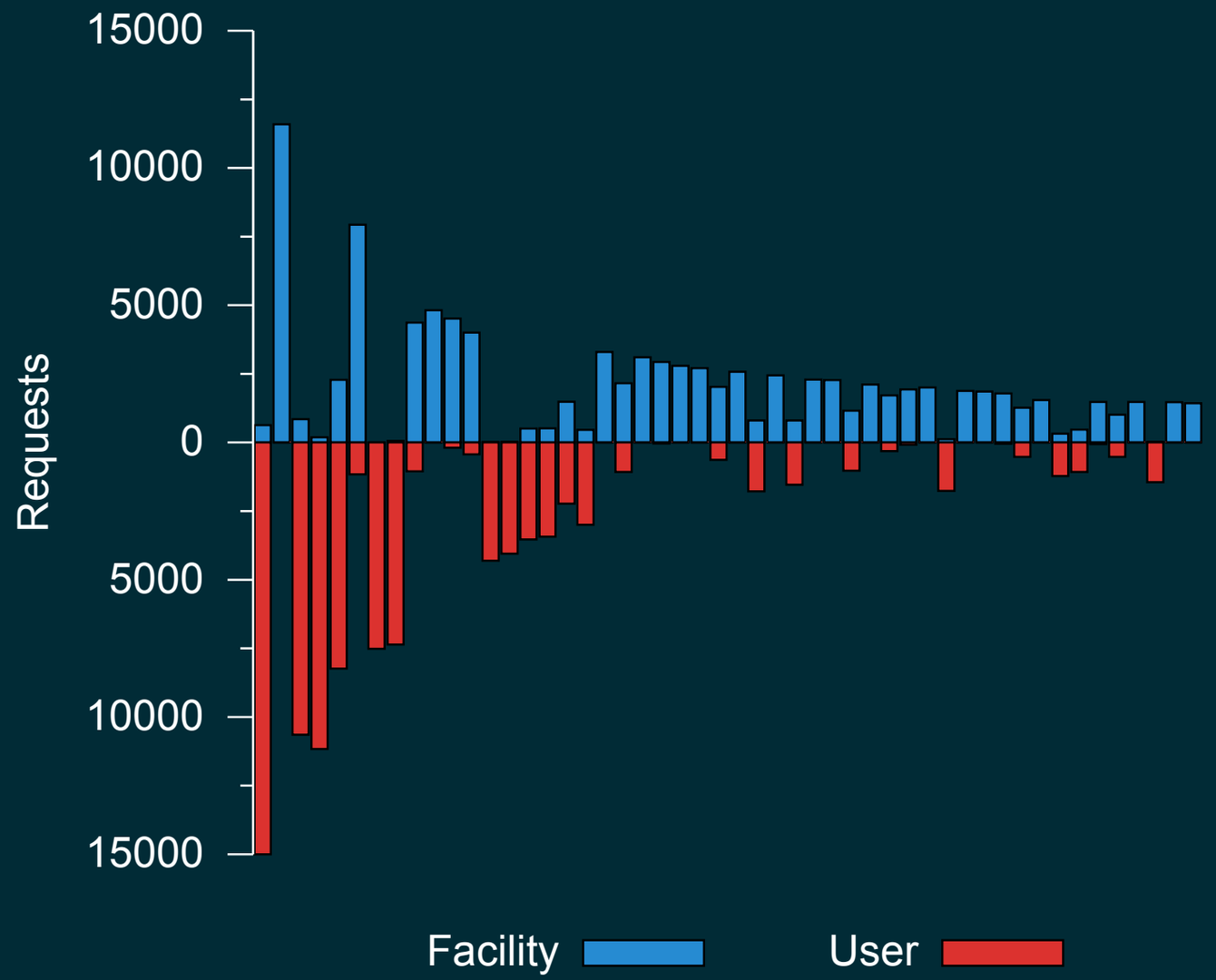
–Reviewer D
[Emulab user]

Facility vs. user images

Facility	User
55.6%	44.4%

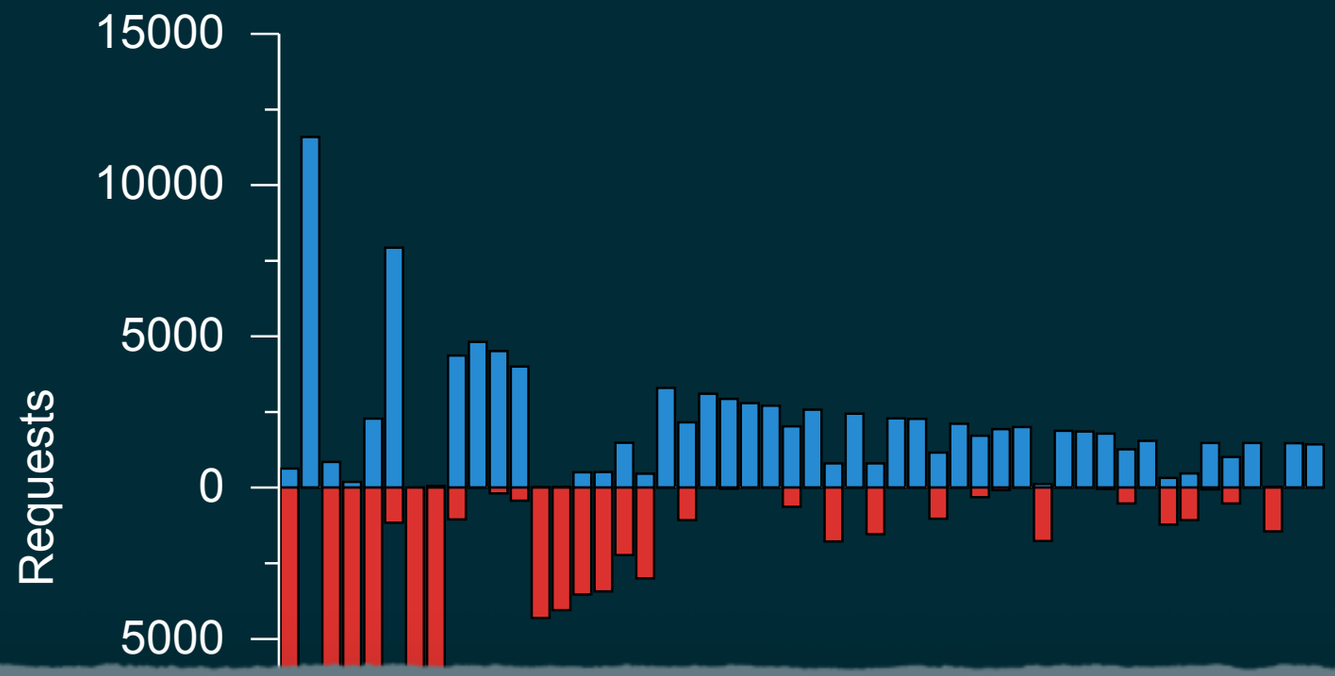
Facility vs. user images

Facility	User
55.6%	44.4%



Facility vs. user images

Facility	User
55.6%	44.4%



1) Most users stick to facility or user images
2) Heaviest users use their own images

Image popularity

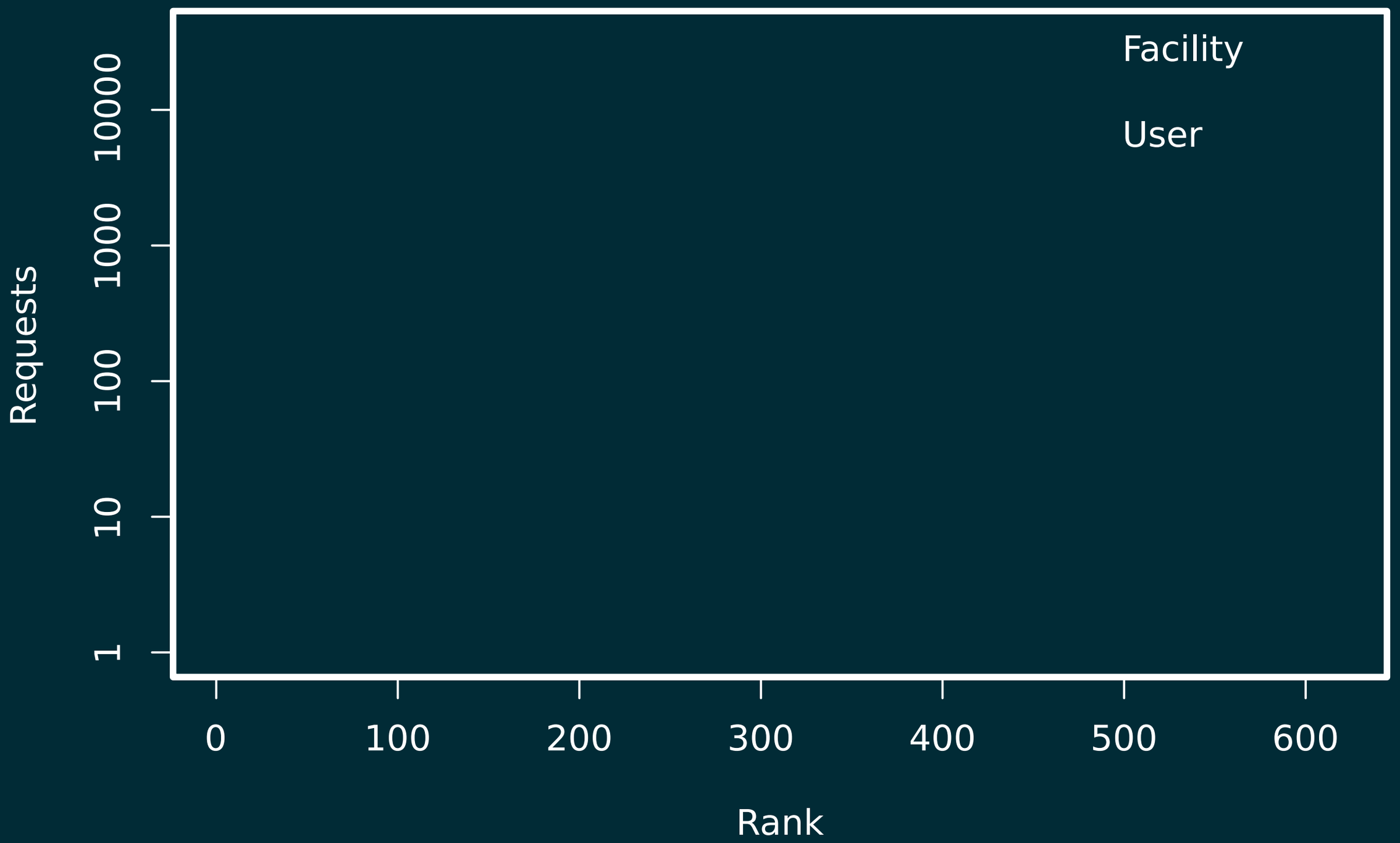


Image popularity

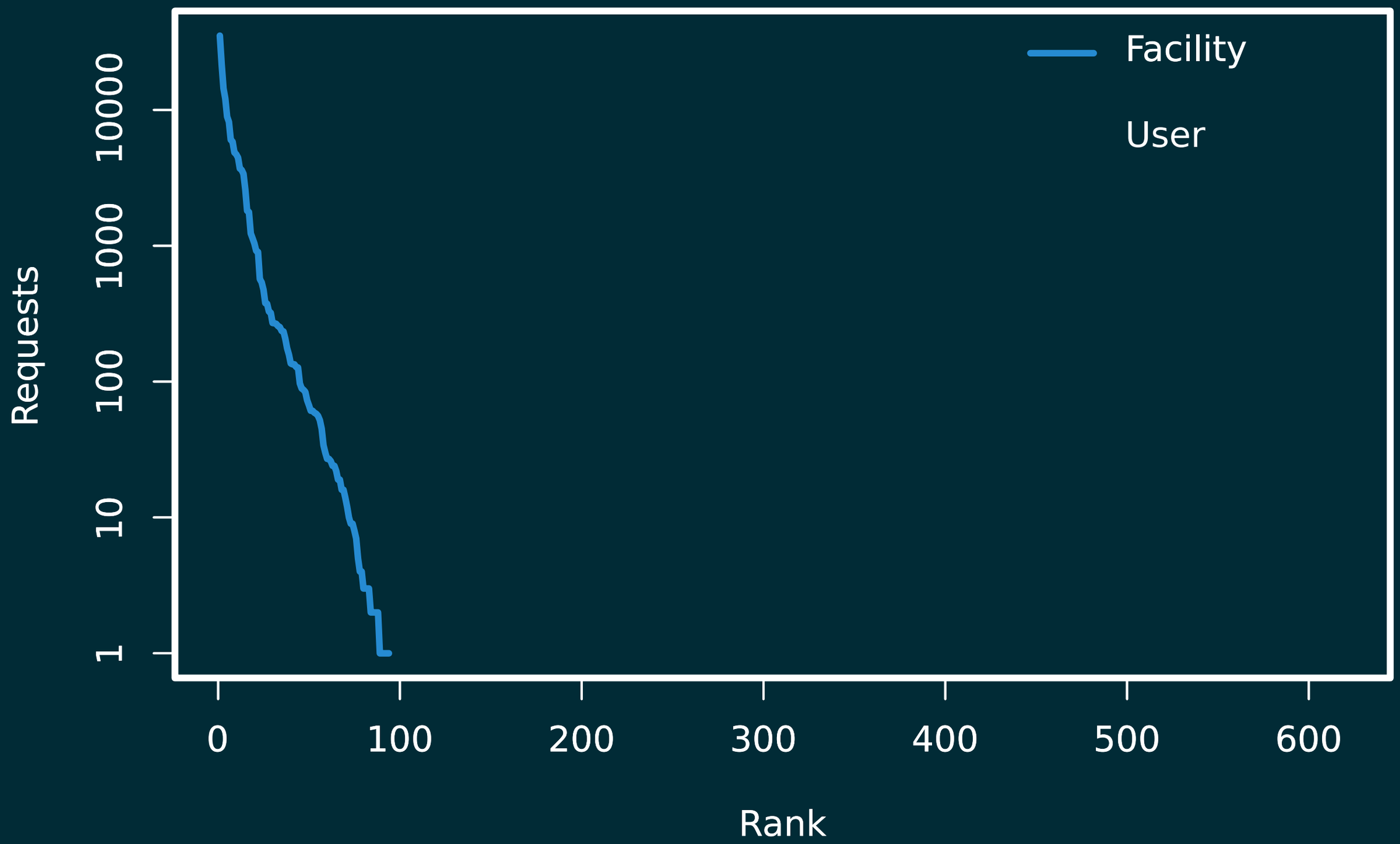


Image popularity

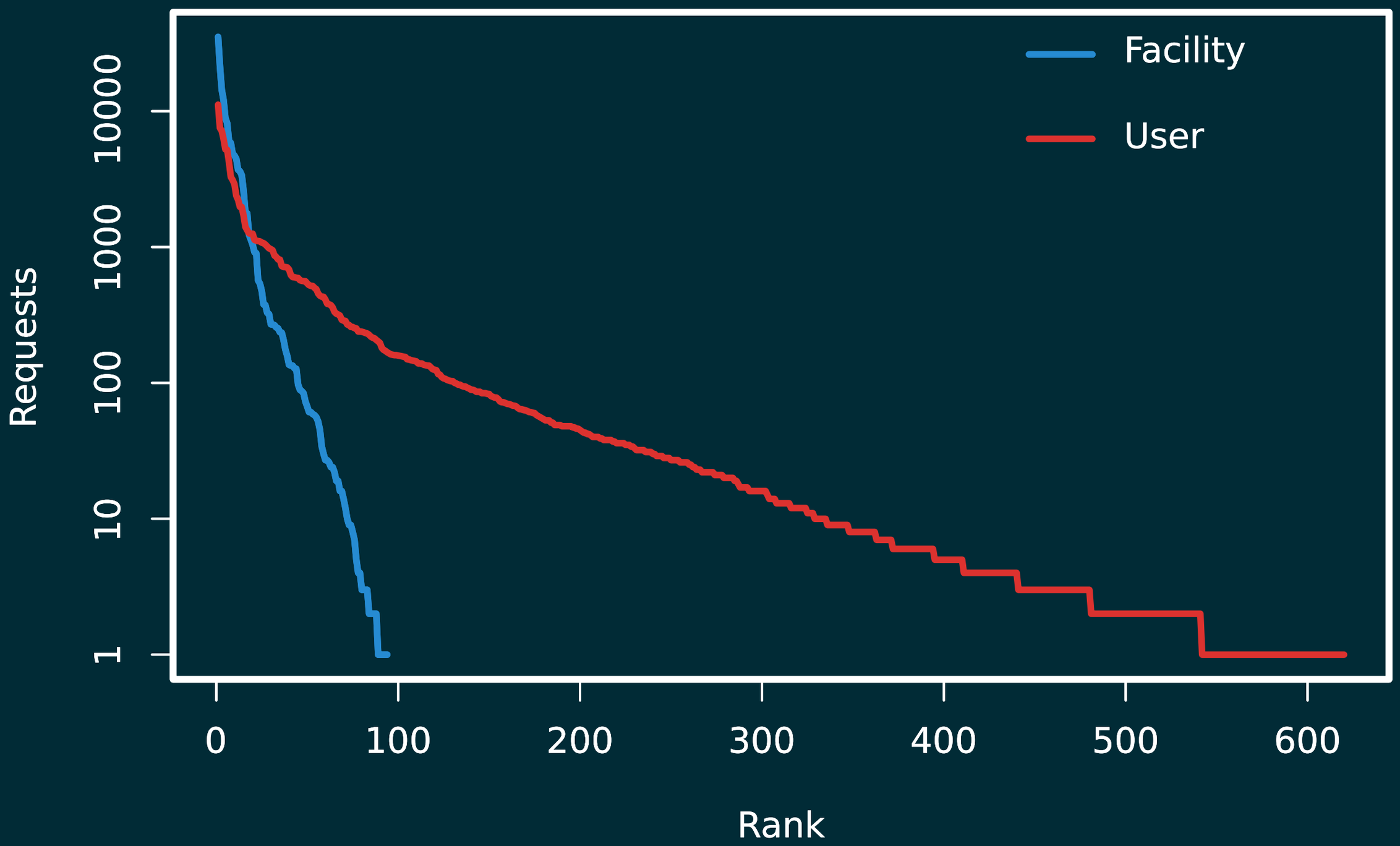


Image popularity

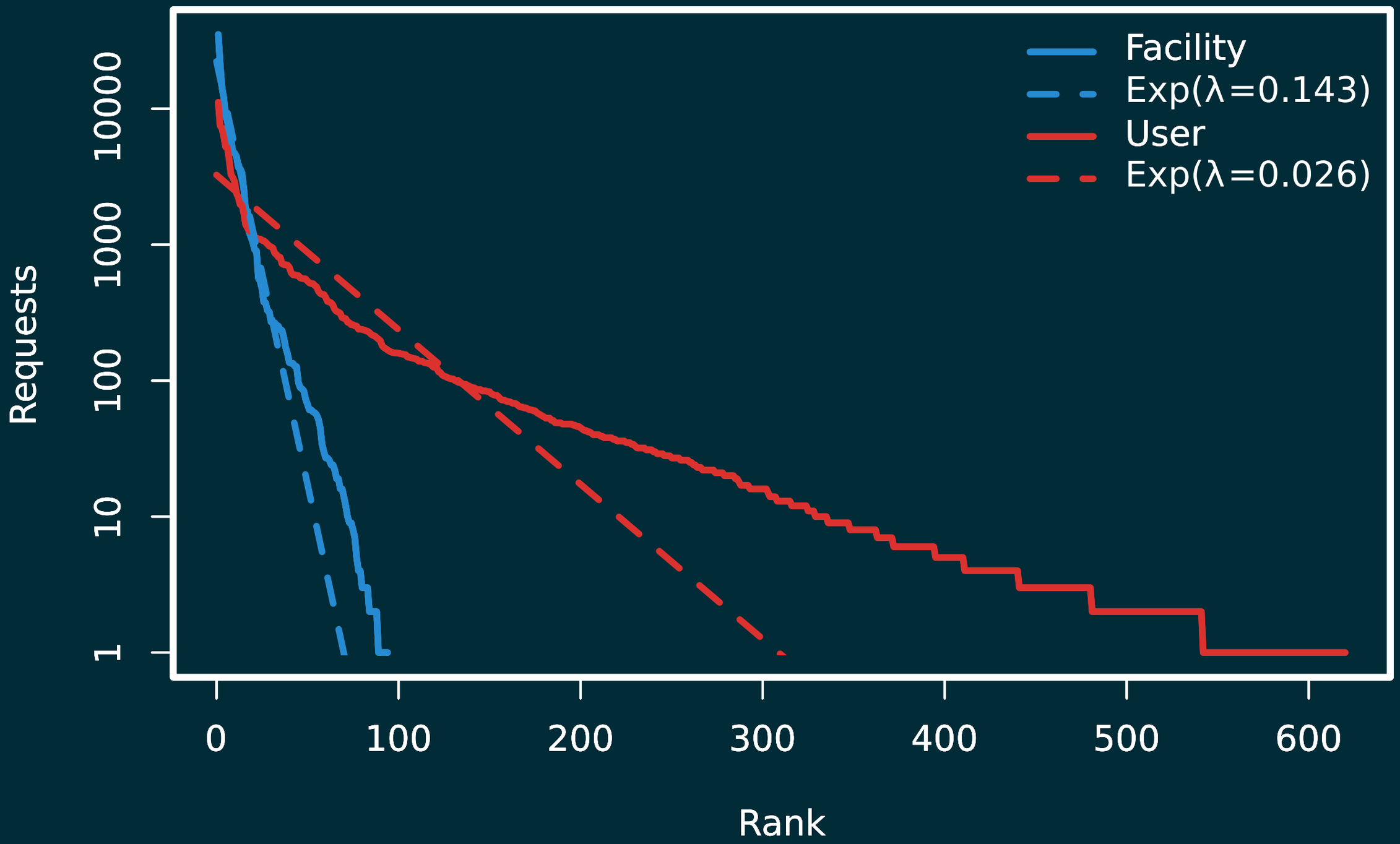


Image popularity

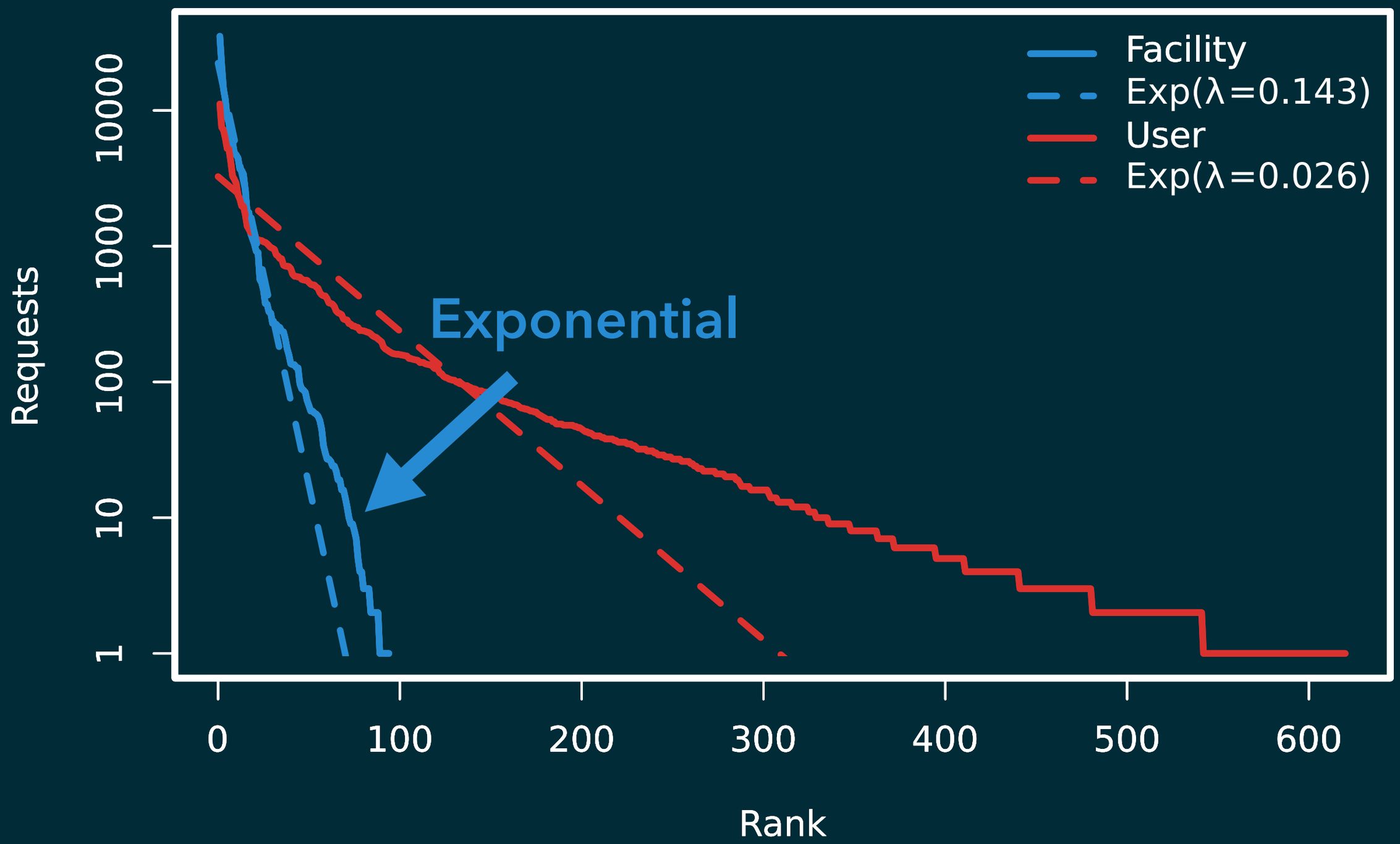


Image popularity

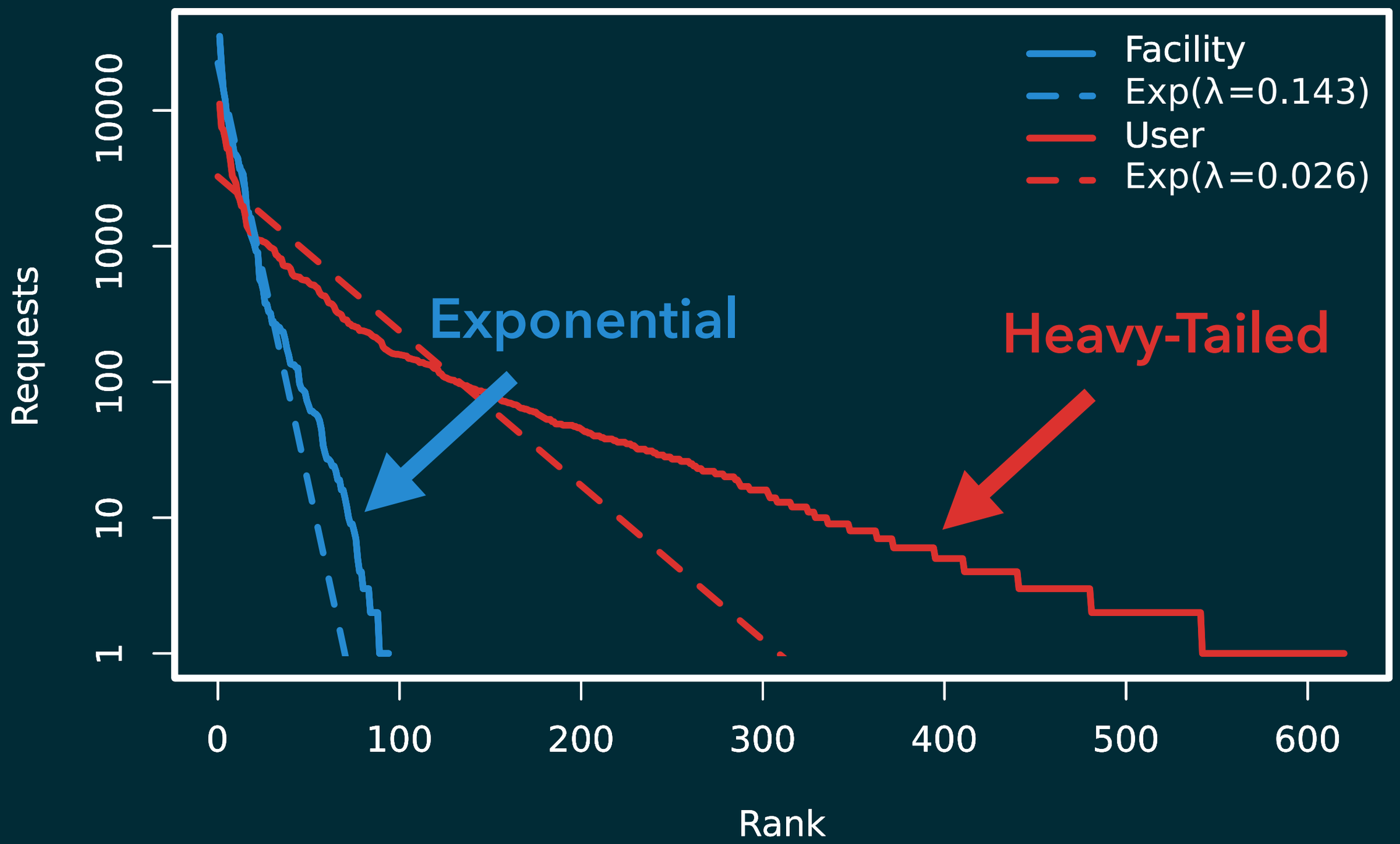
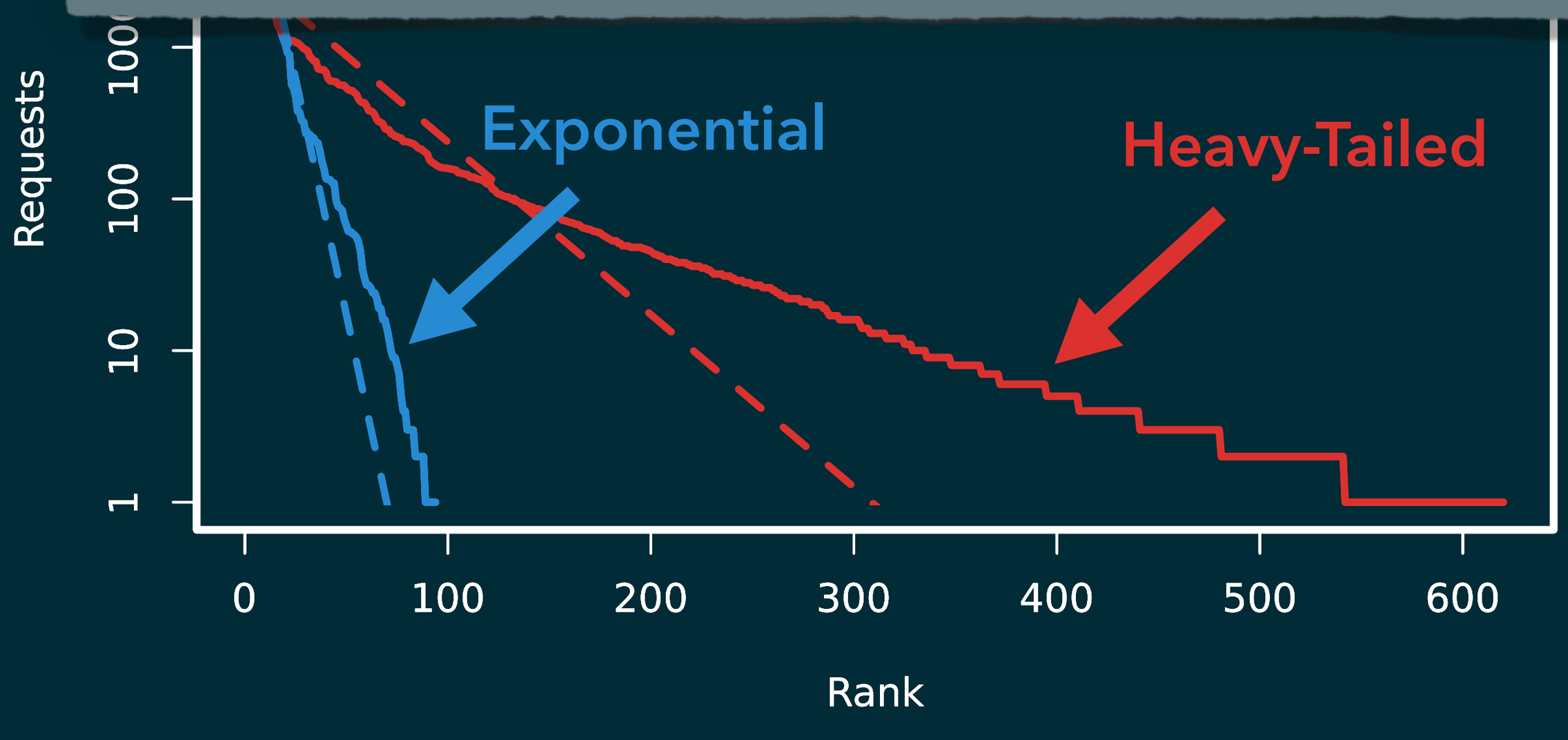
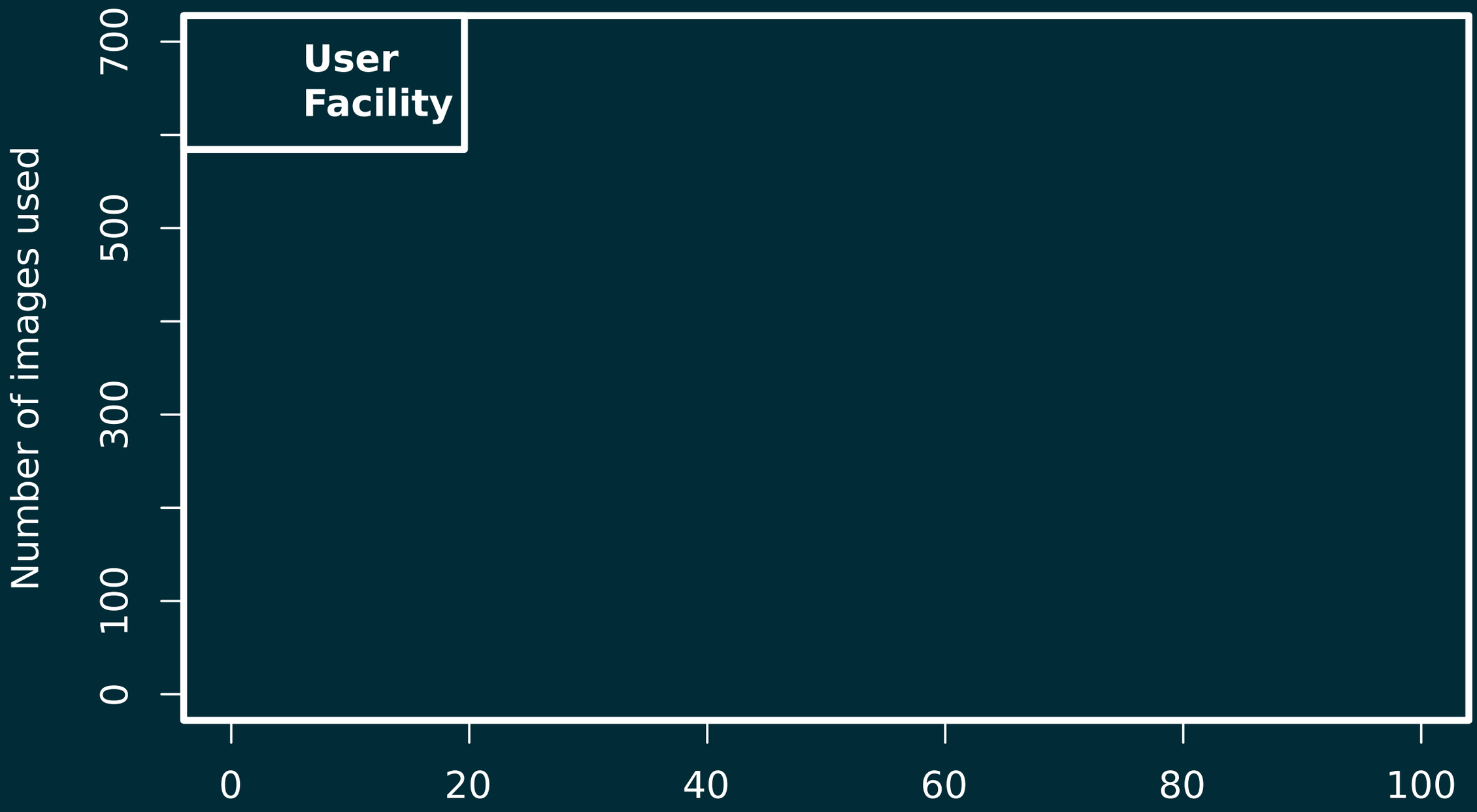


Image popularity

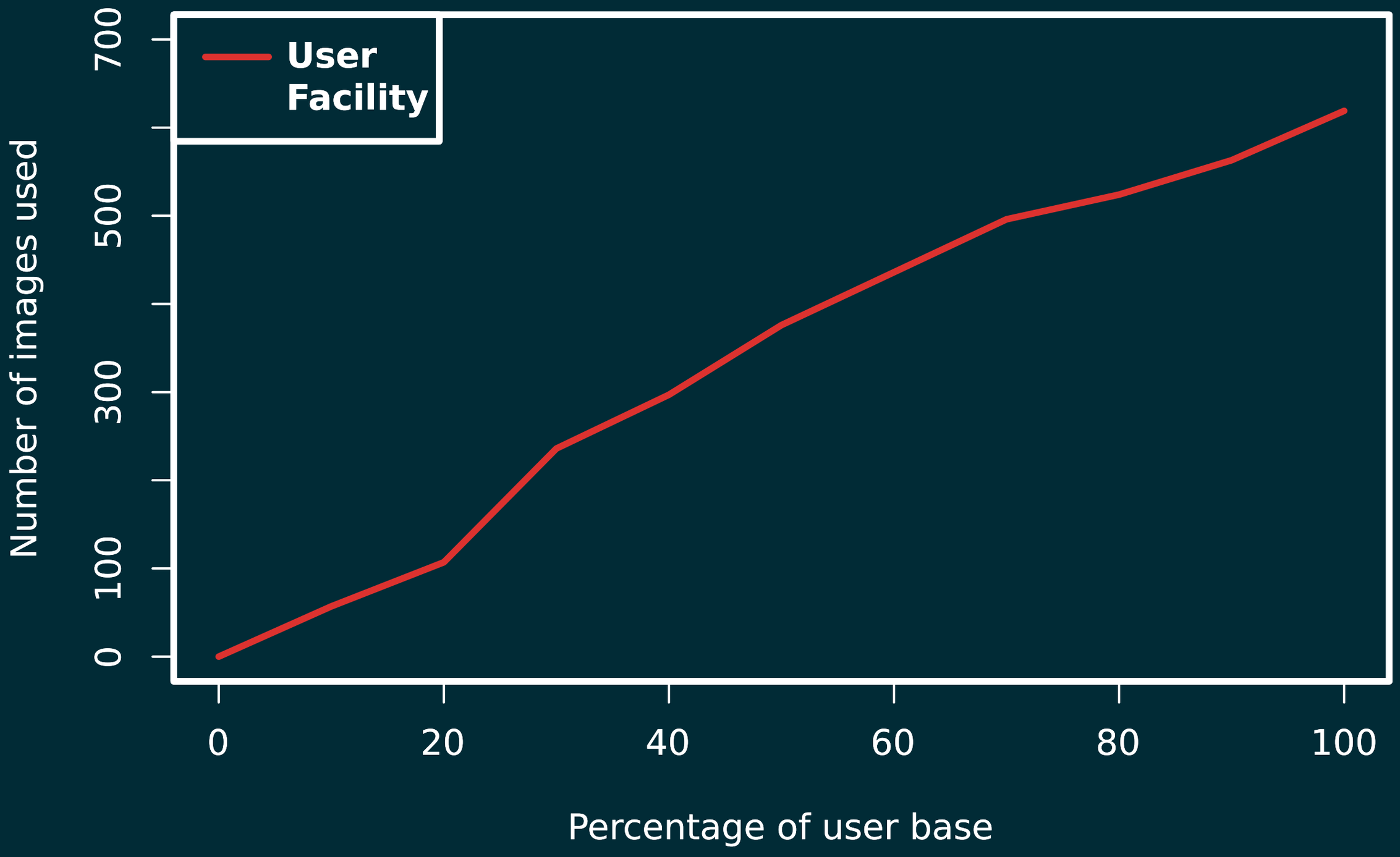
1) Facility images have a smaller, lighter tail
2) Most popular image < 13% of requests



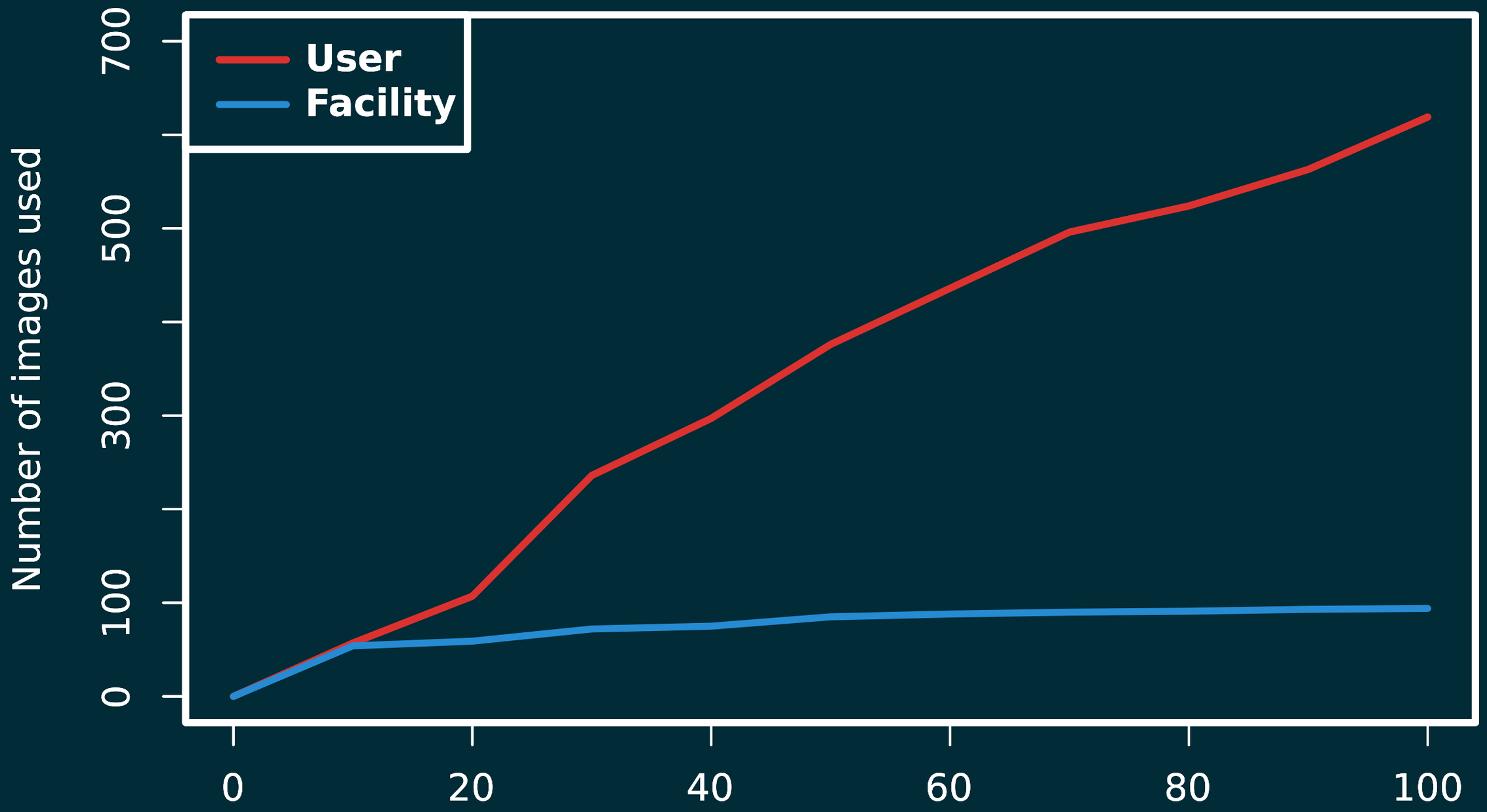
Scaling: total images



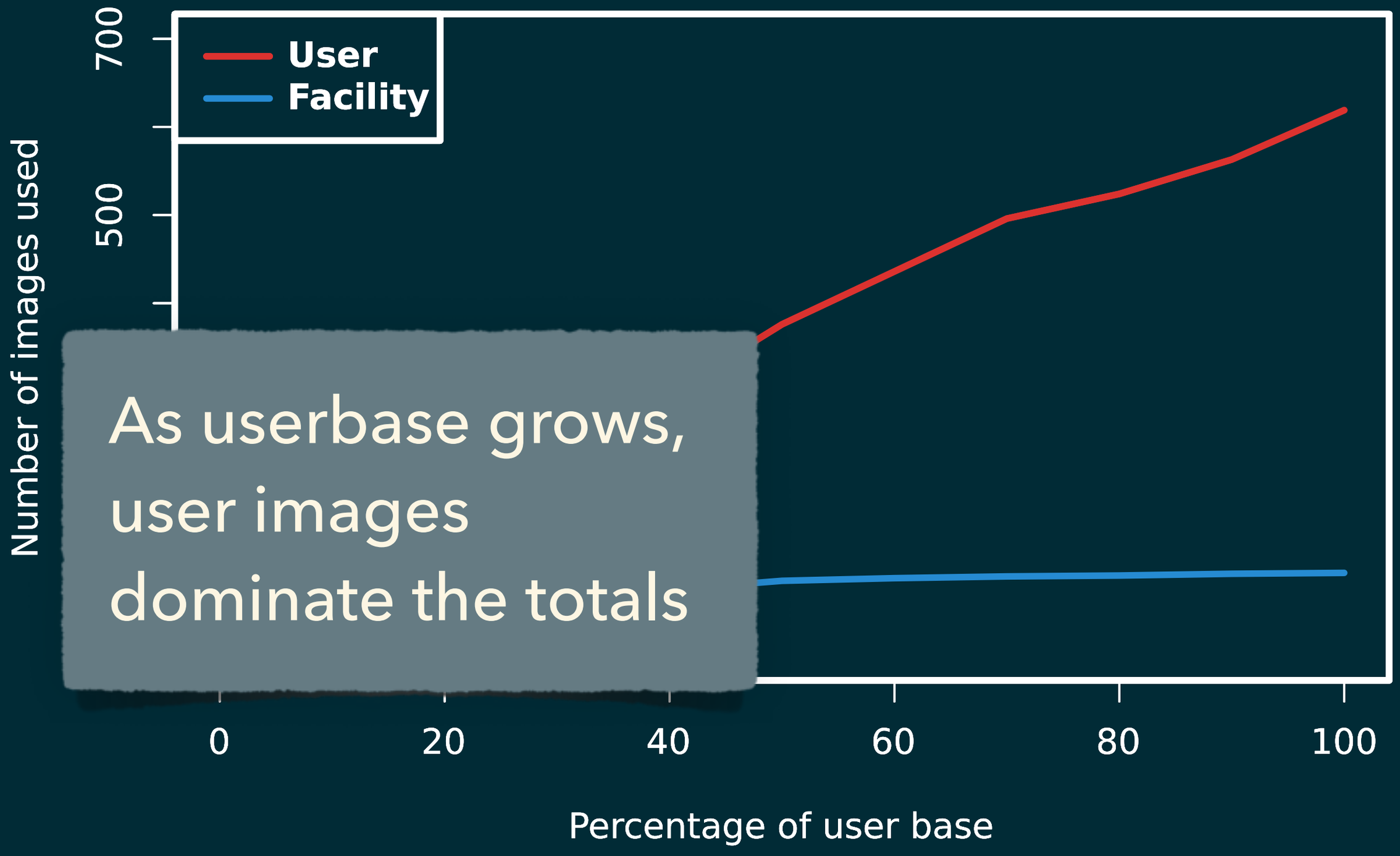
Scaling: total images



Scaling: total images

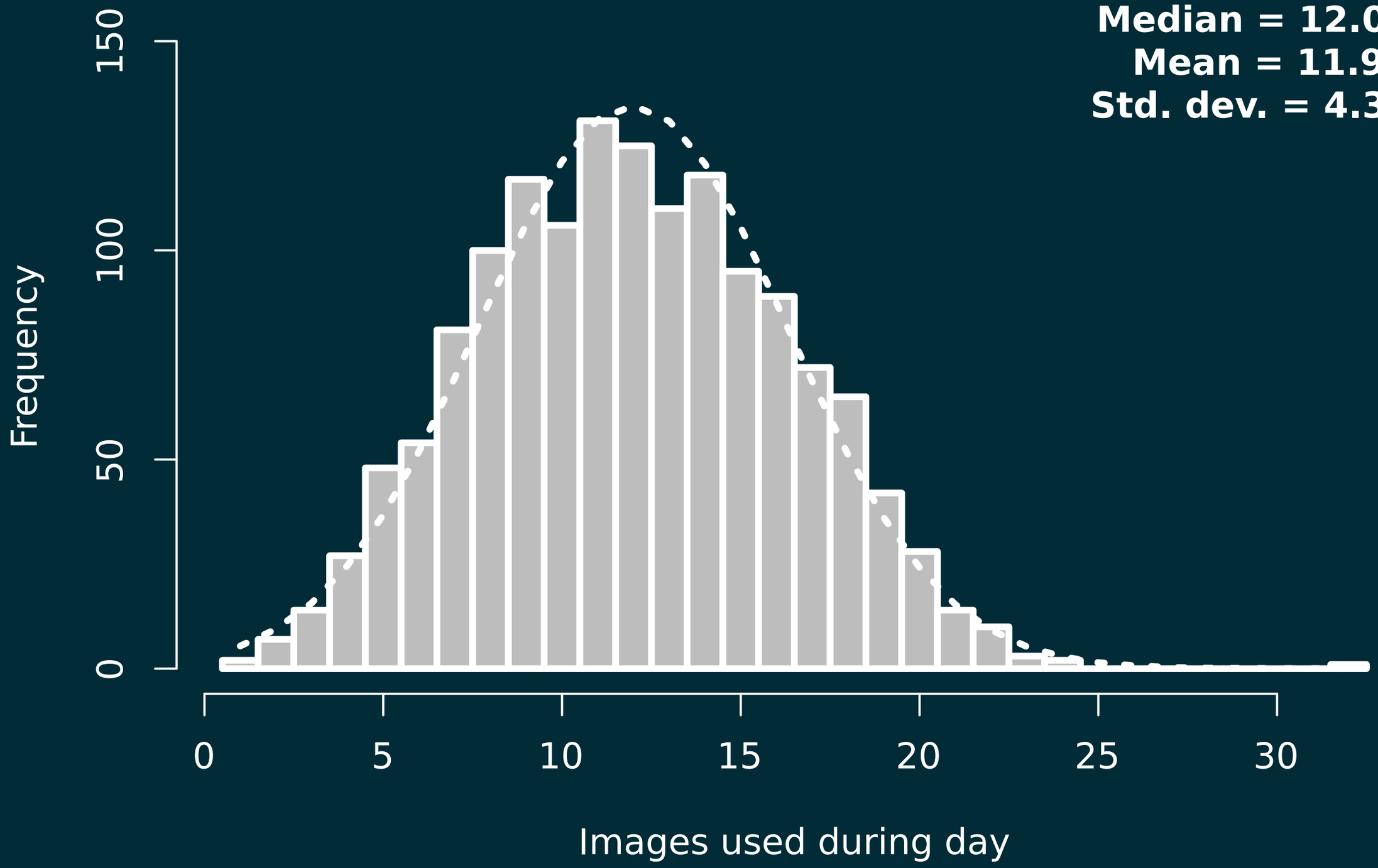


Scaling: total images

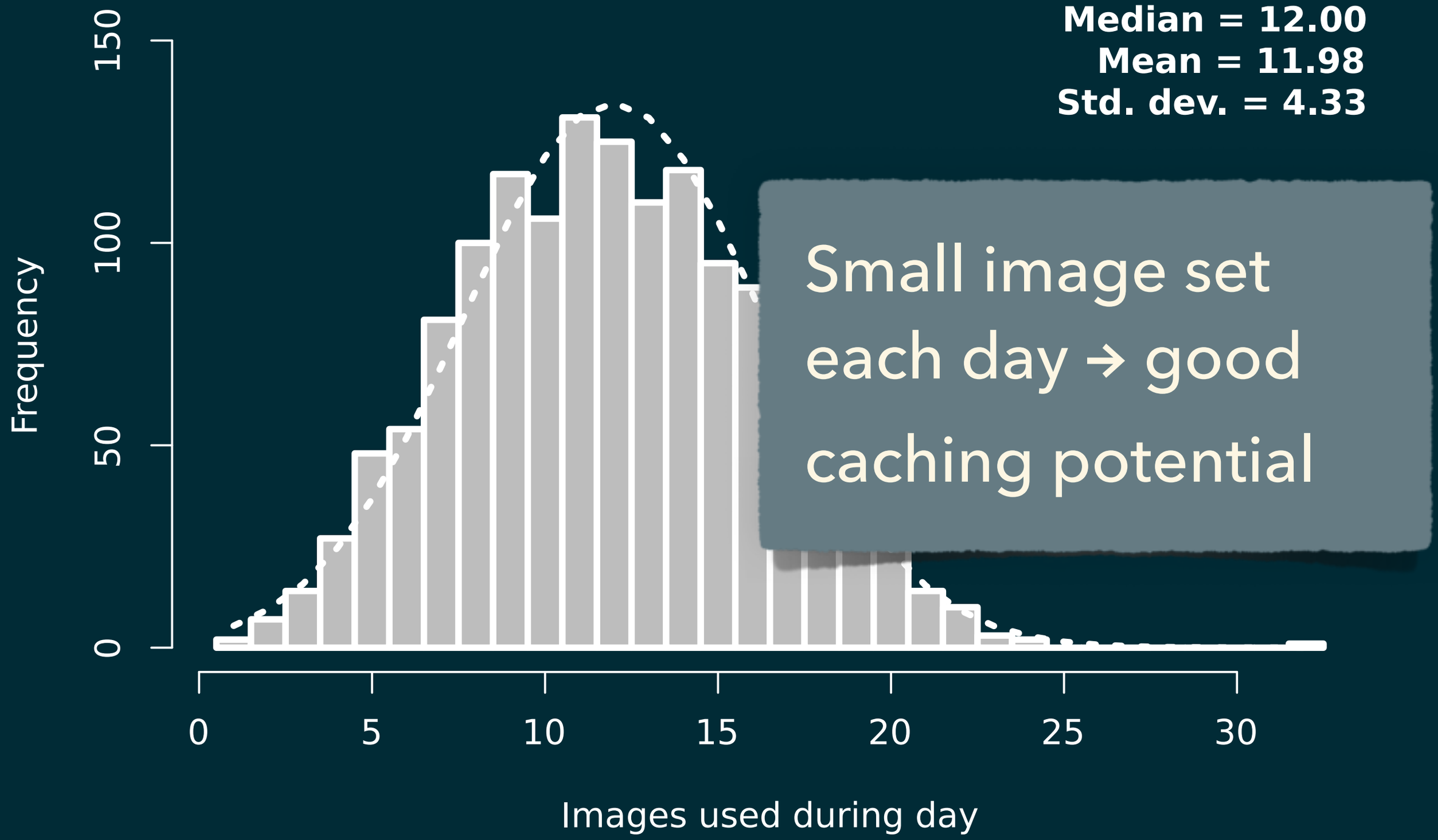


As userbase grows, user images dominate the totals

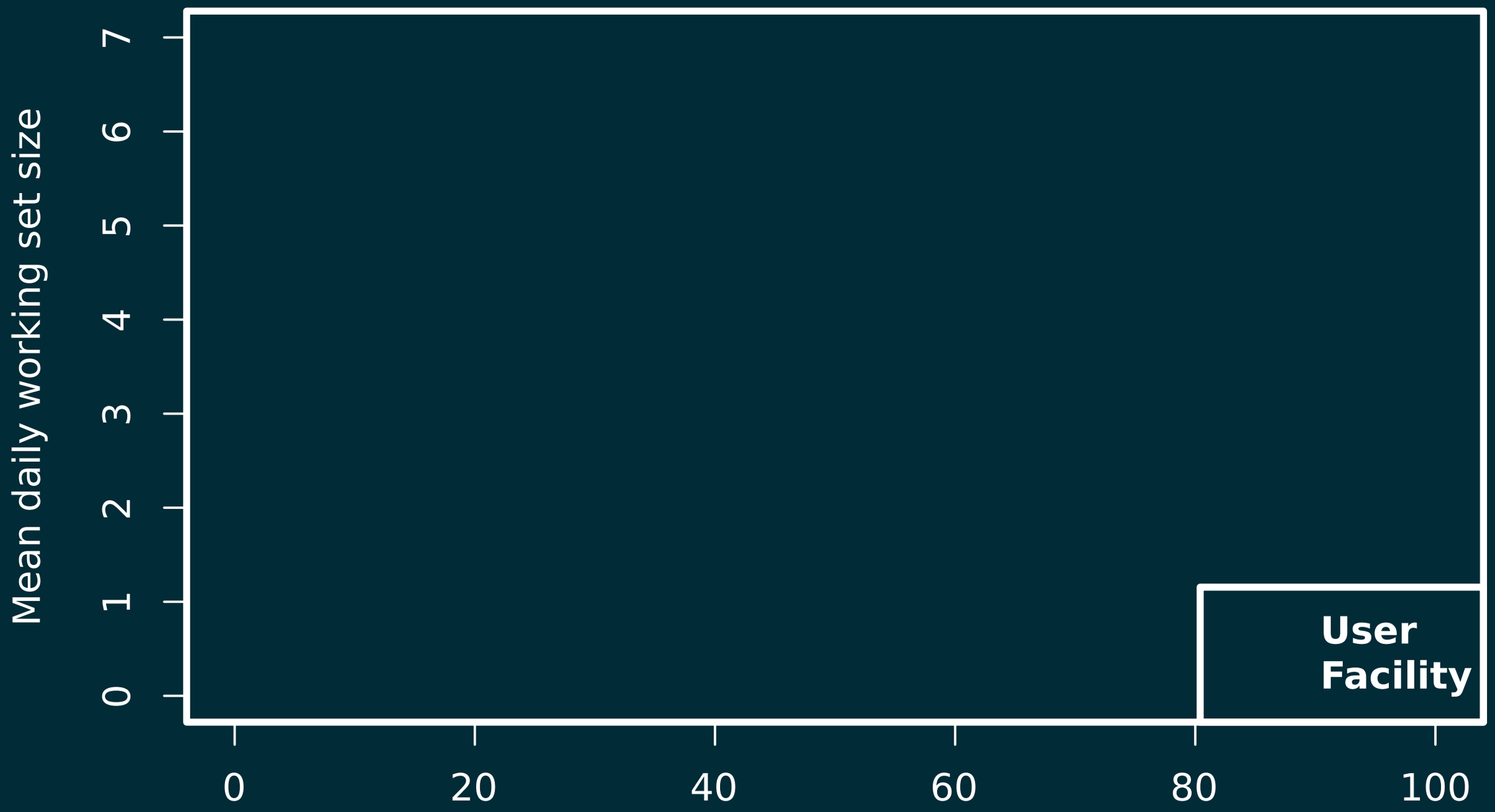
Daily working set



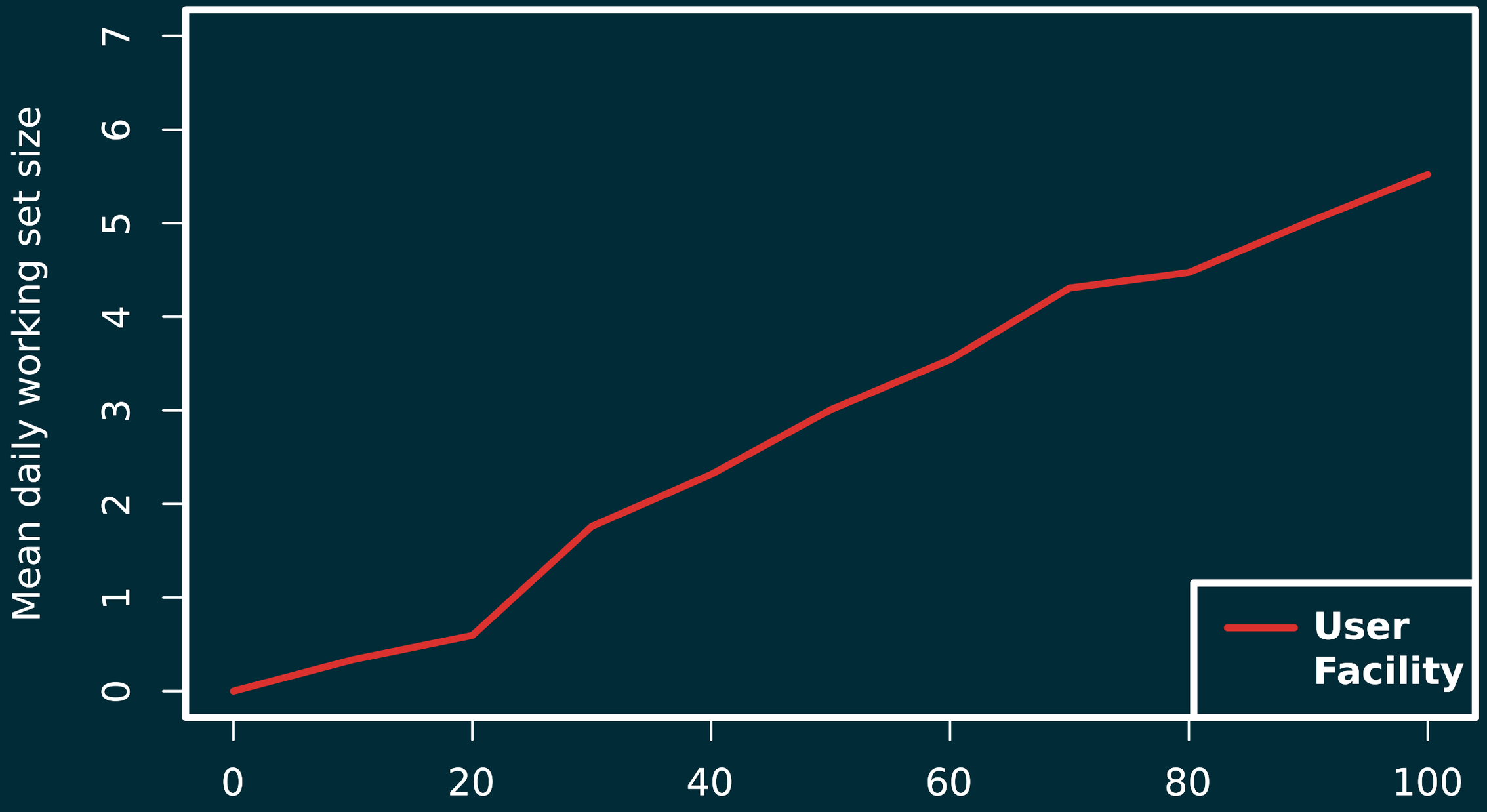
Daily working set



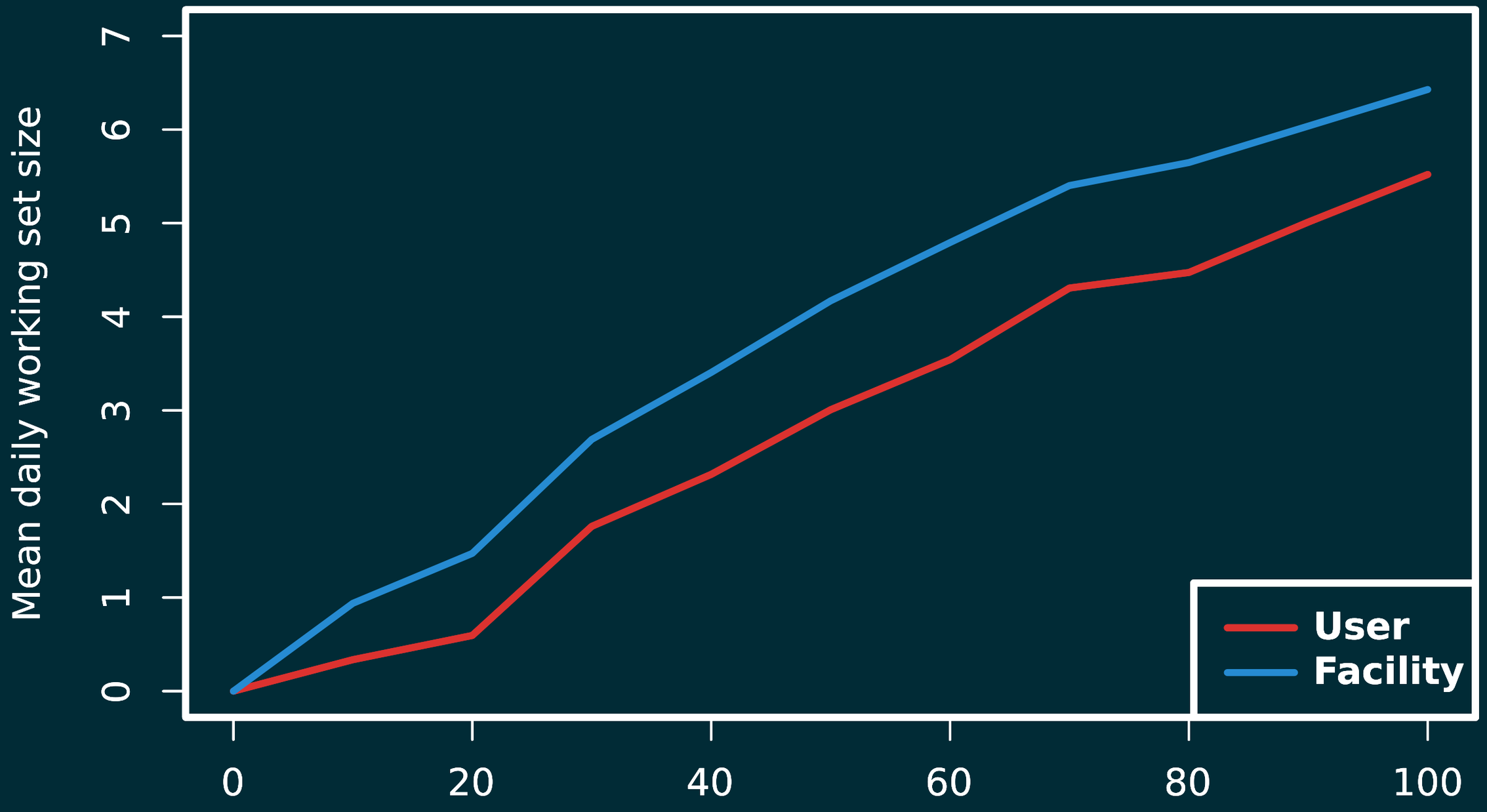
Scaling: working set



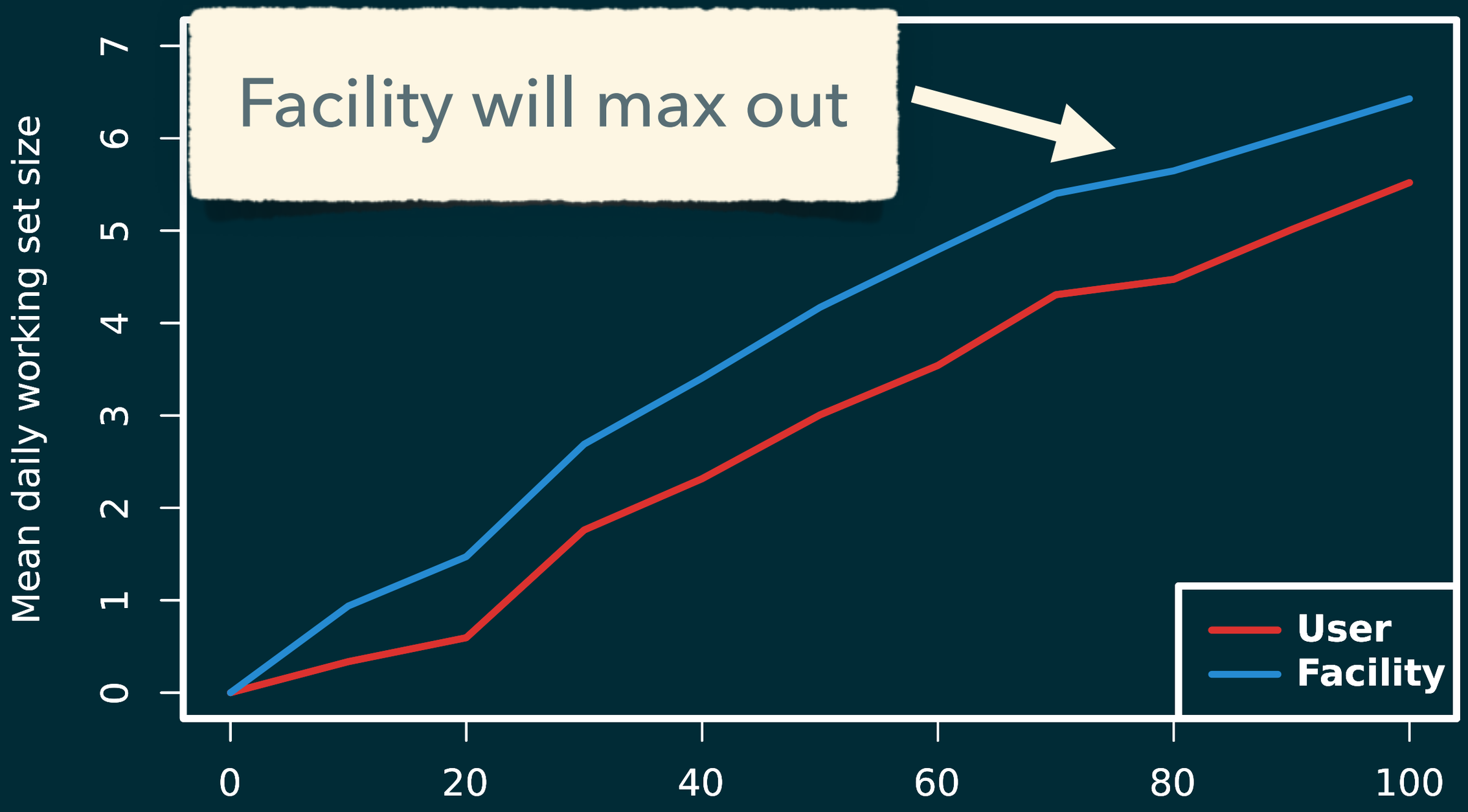
Scaling: working set



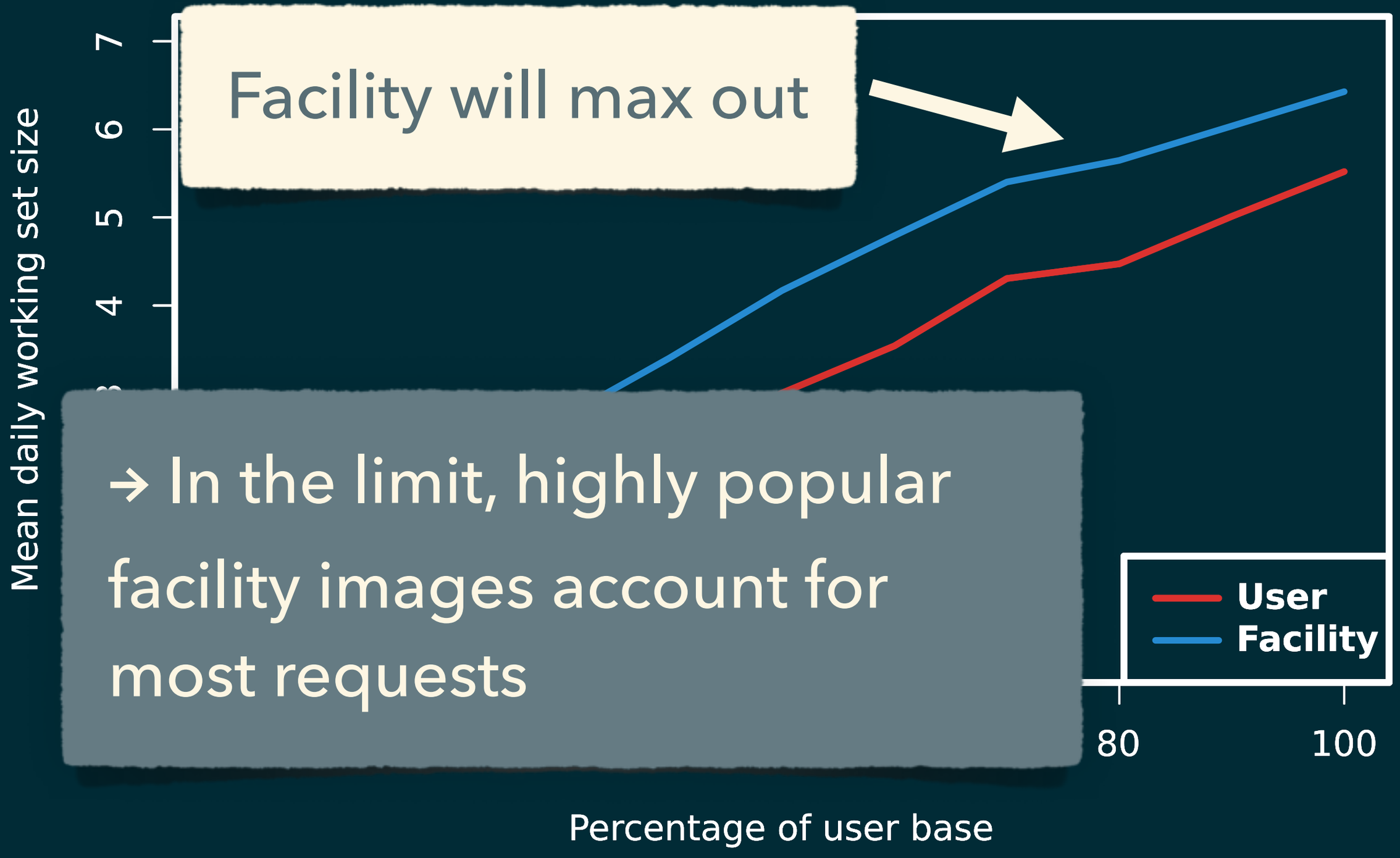
Scaling: working set



Scaling: working set



Scaling: working set



→ In the limit, highly popular facility images account for most requests

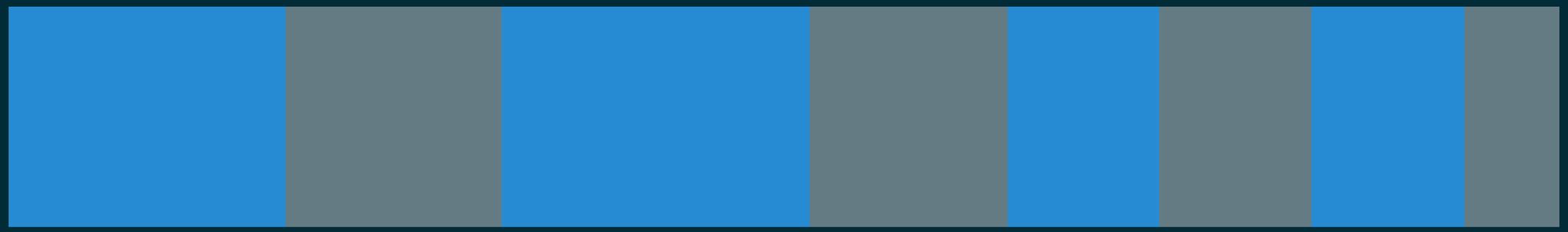
Image Contents

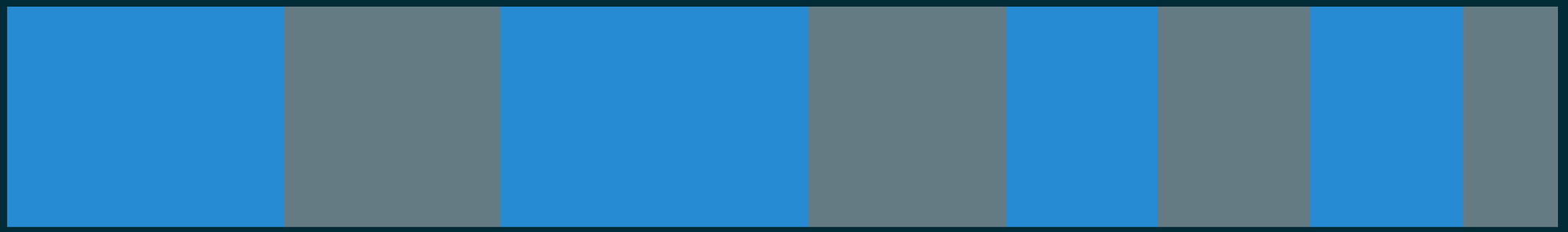
Block-level similarity

Base

Block-level similarity

Base

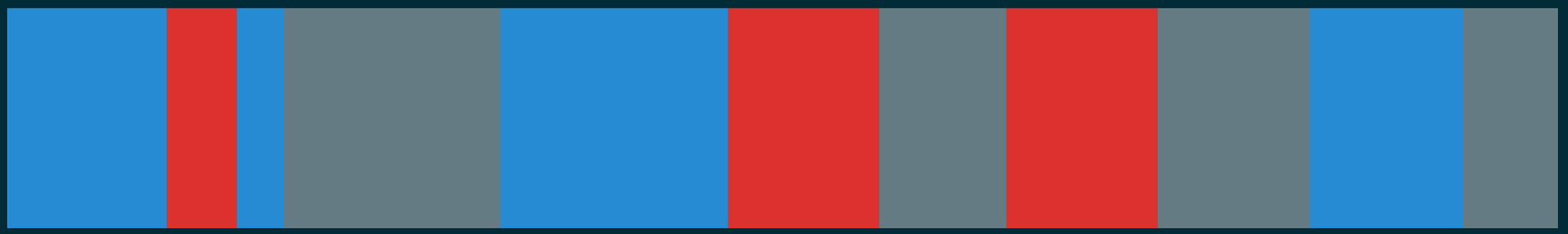




Derived

Block-level similarity

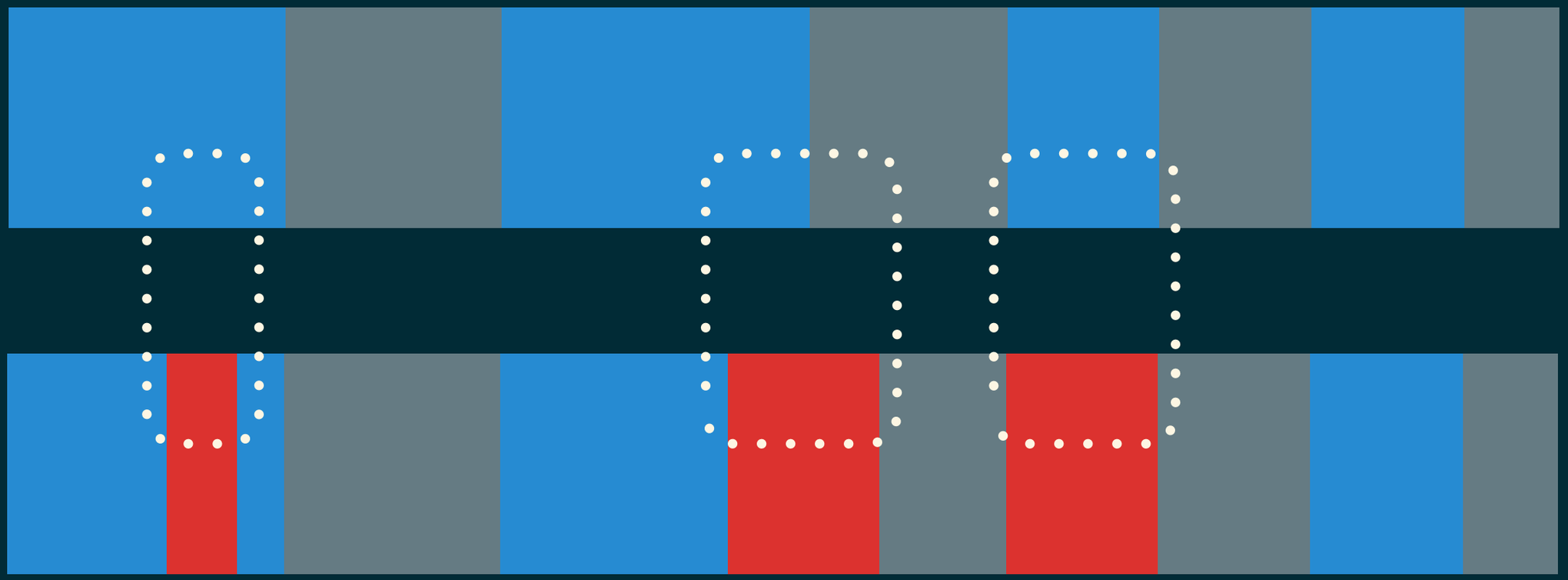
Base



Derived

Block-level similarity

Base



Derived

Percentage of blocks that need to be written to transform the base image into derived

Block-level similarity

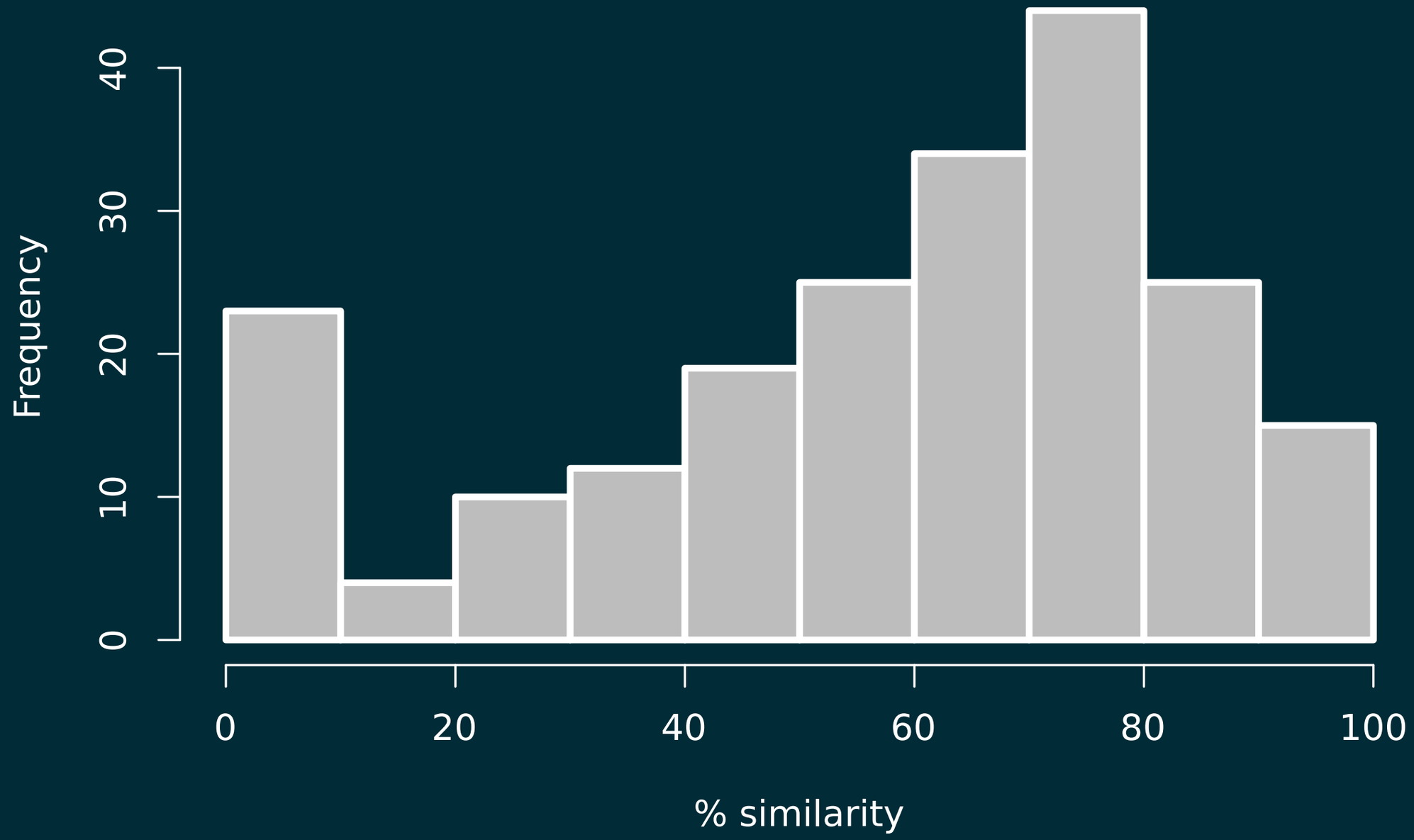
Derived: User image

Base: Most similar facility image

Block-level similarity

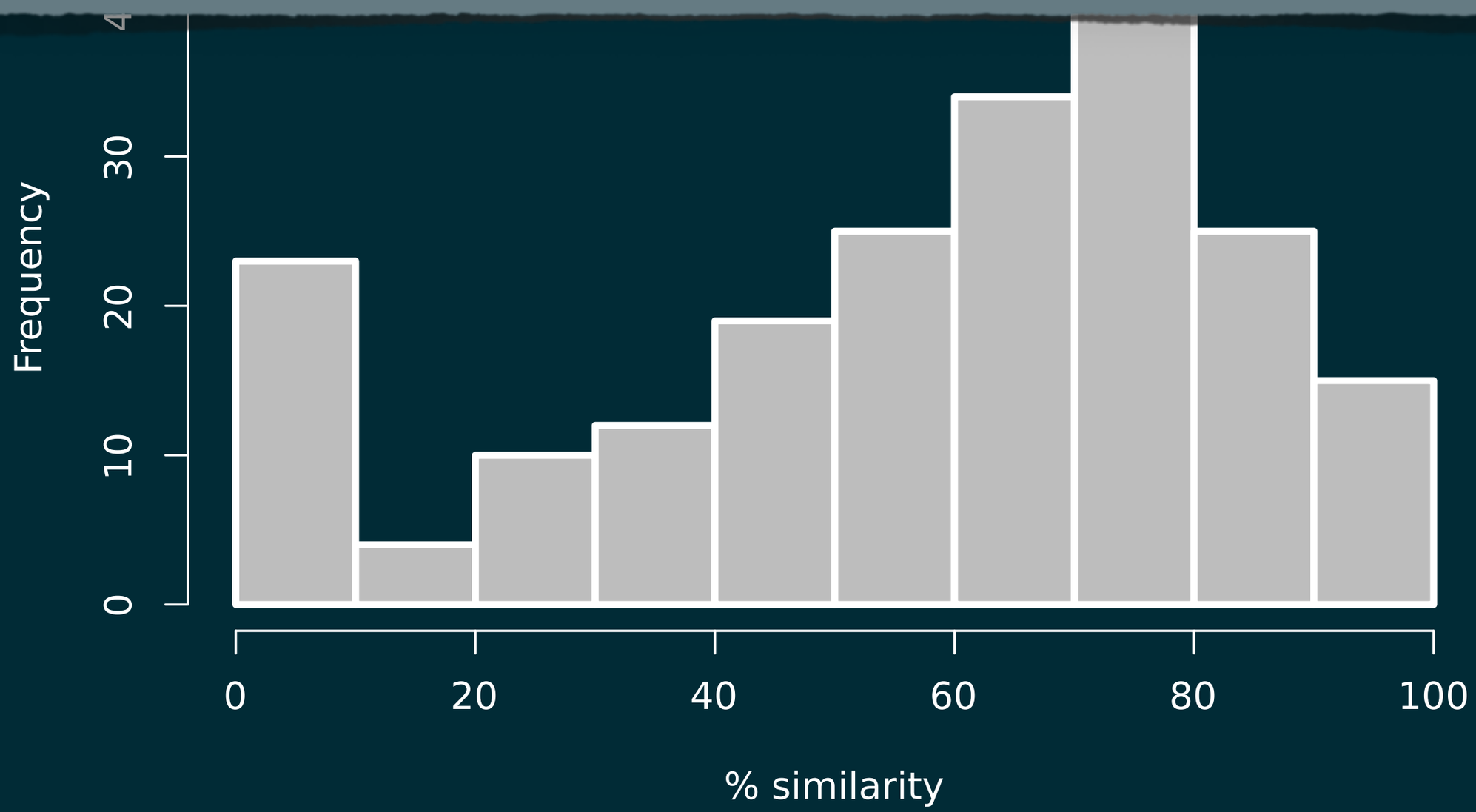
Derived: User image

Base: Most similar facility image



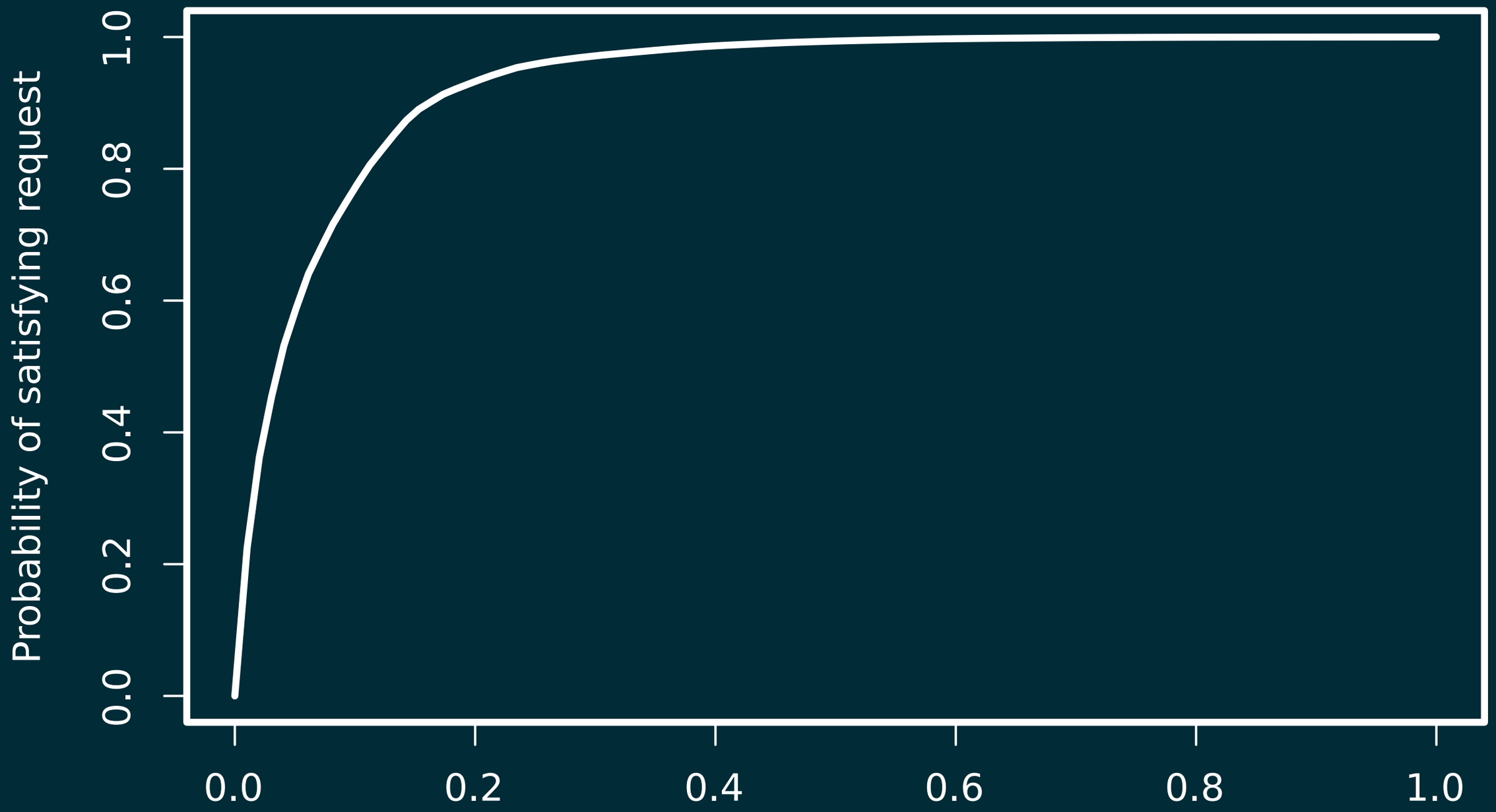
Block-level similarity

1) De-duplicating storage an attractive option
2) Differential loading has potential

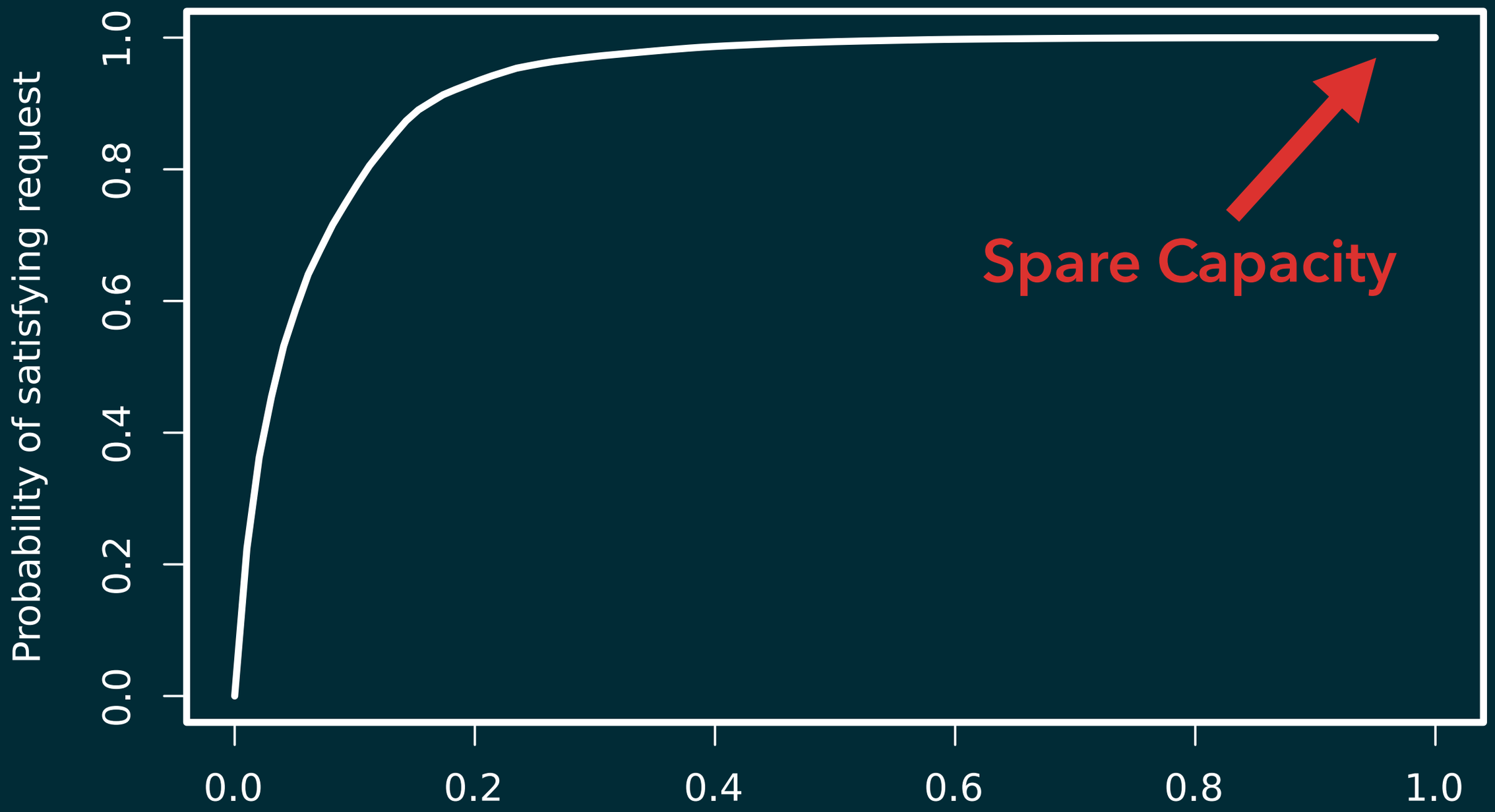


Pre-Loading

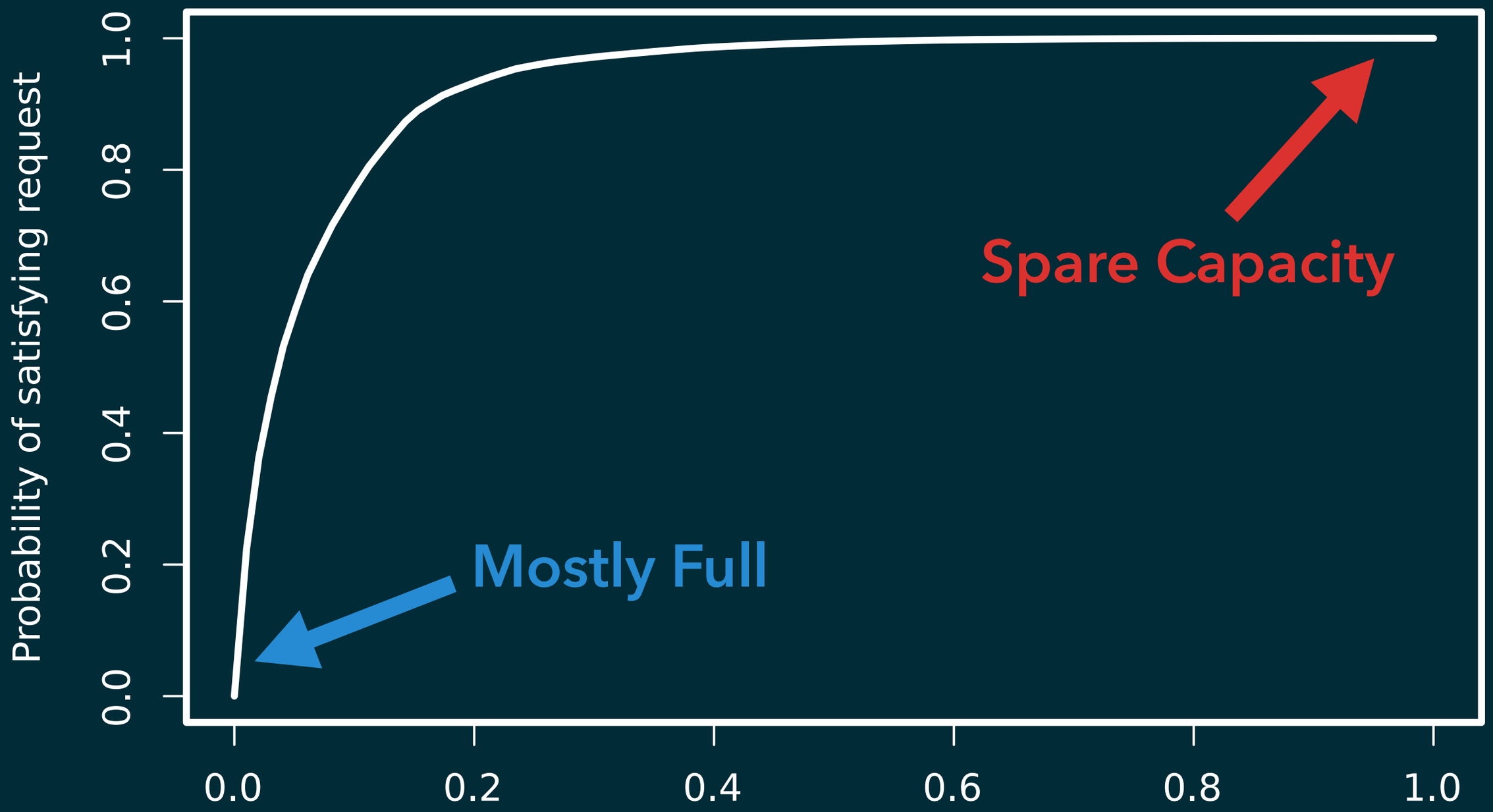
Pre-loading: Size



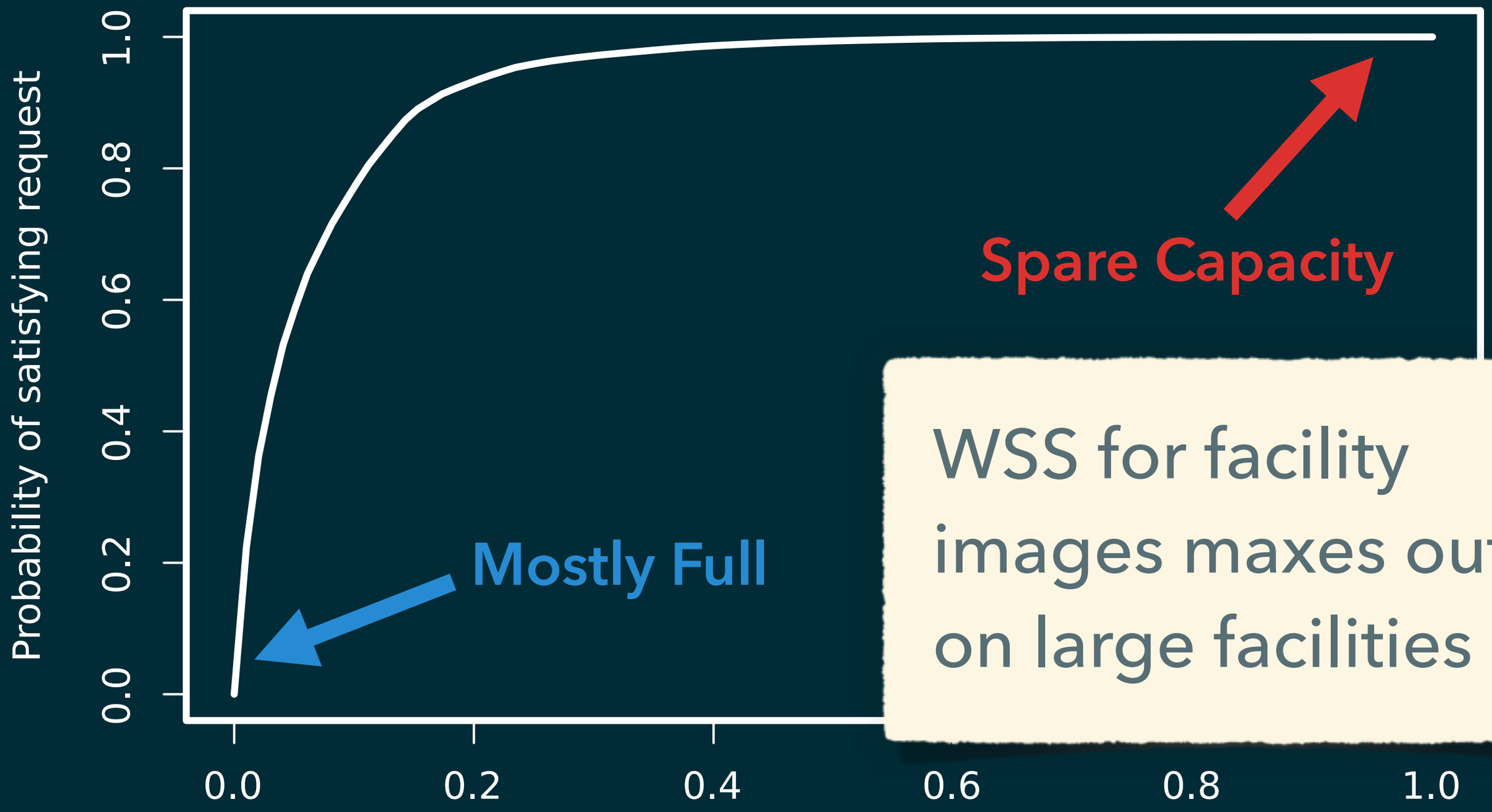
Pre-loading: Size



Pre-loading: Size



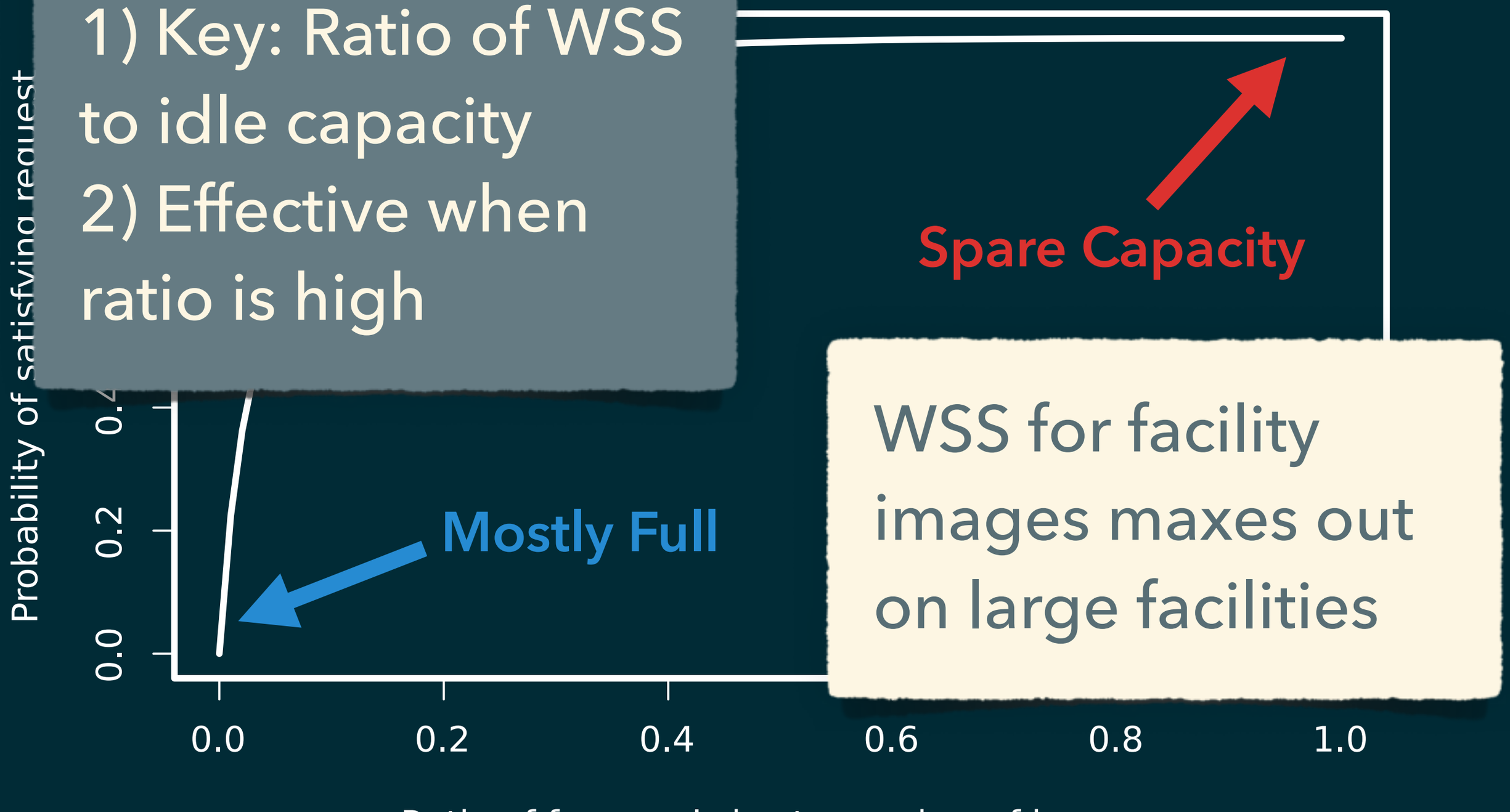
Pre-loading: Size



WSS for facility images maxes out on large facilities

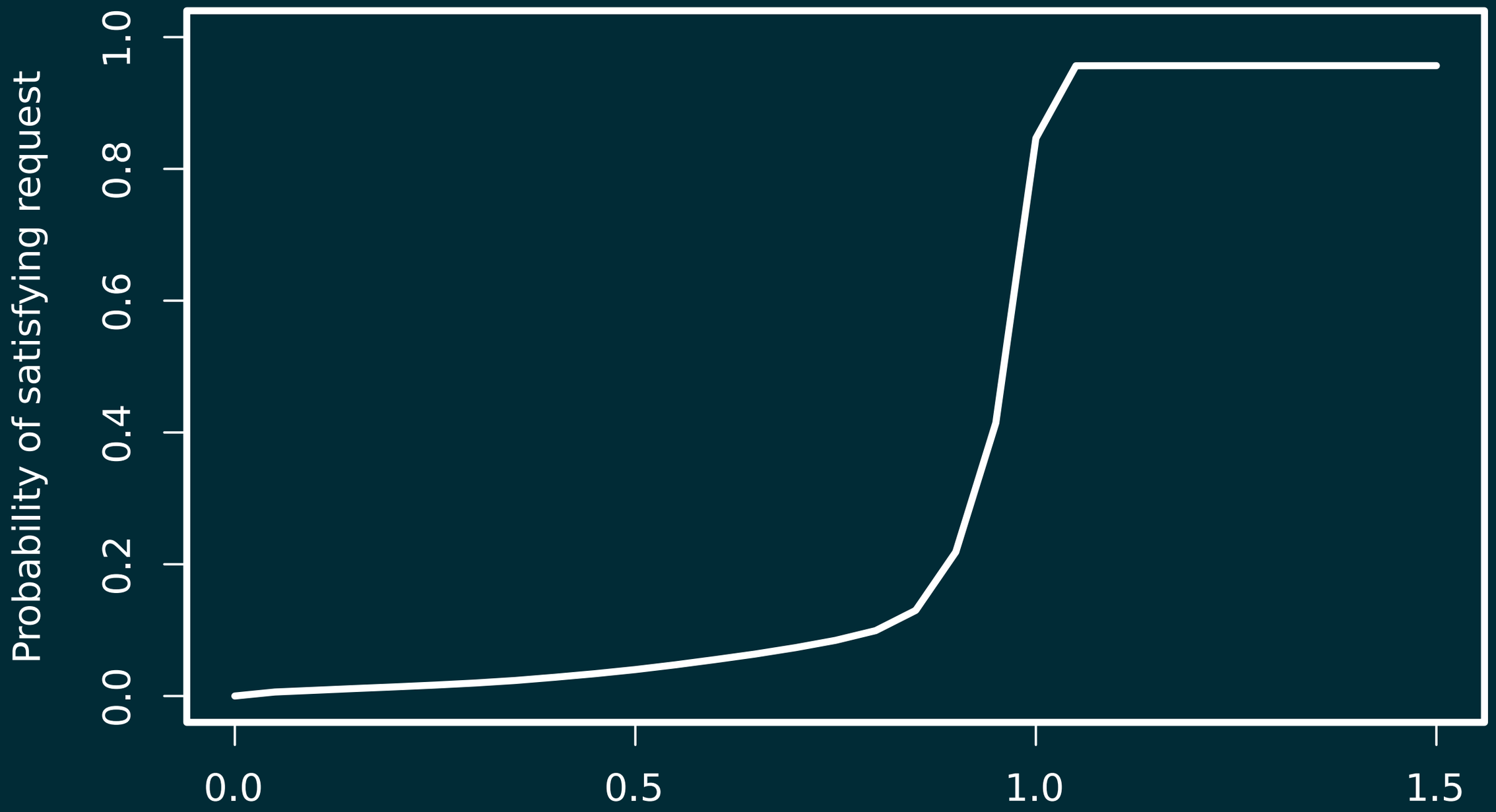
Pre-loading: Size

1) Key: Ratio of WSS to idle capacity
2) Effective when ratio is high

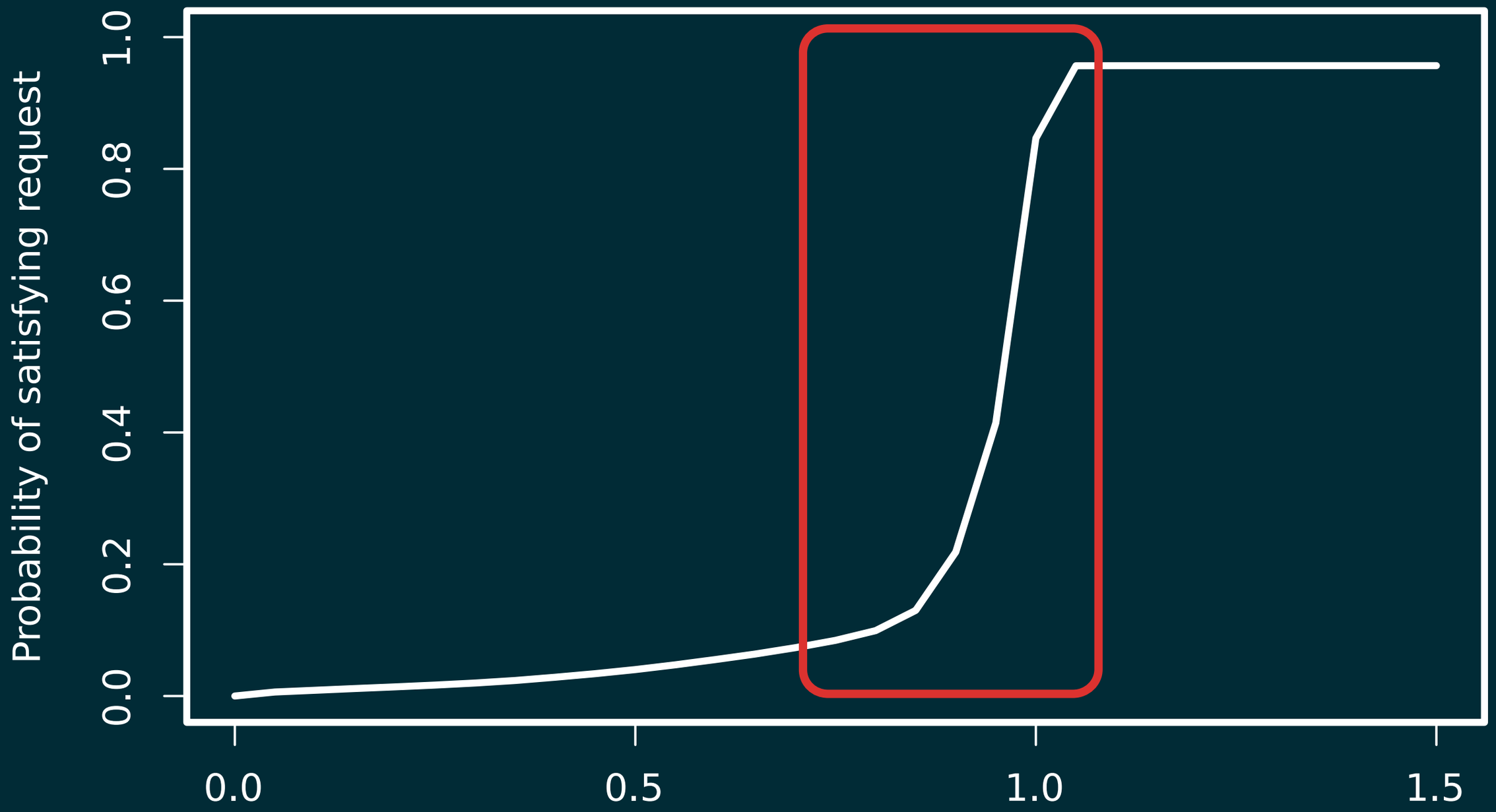


WSS for facility images maxes out on large facilities

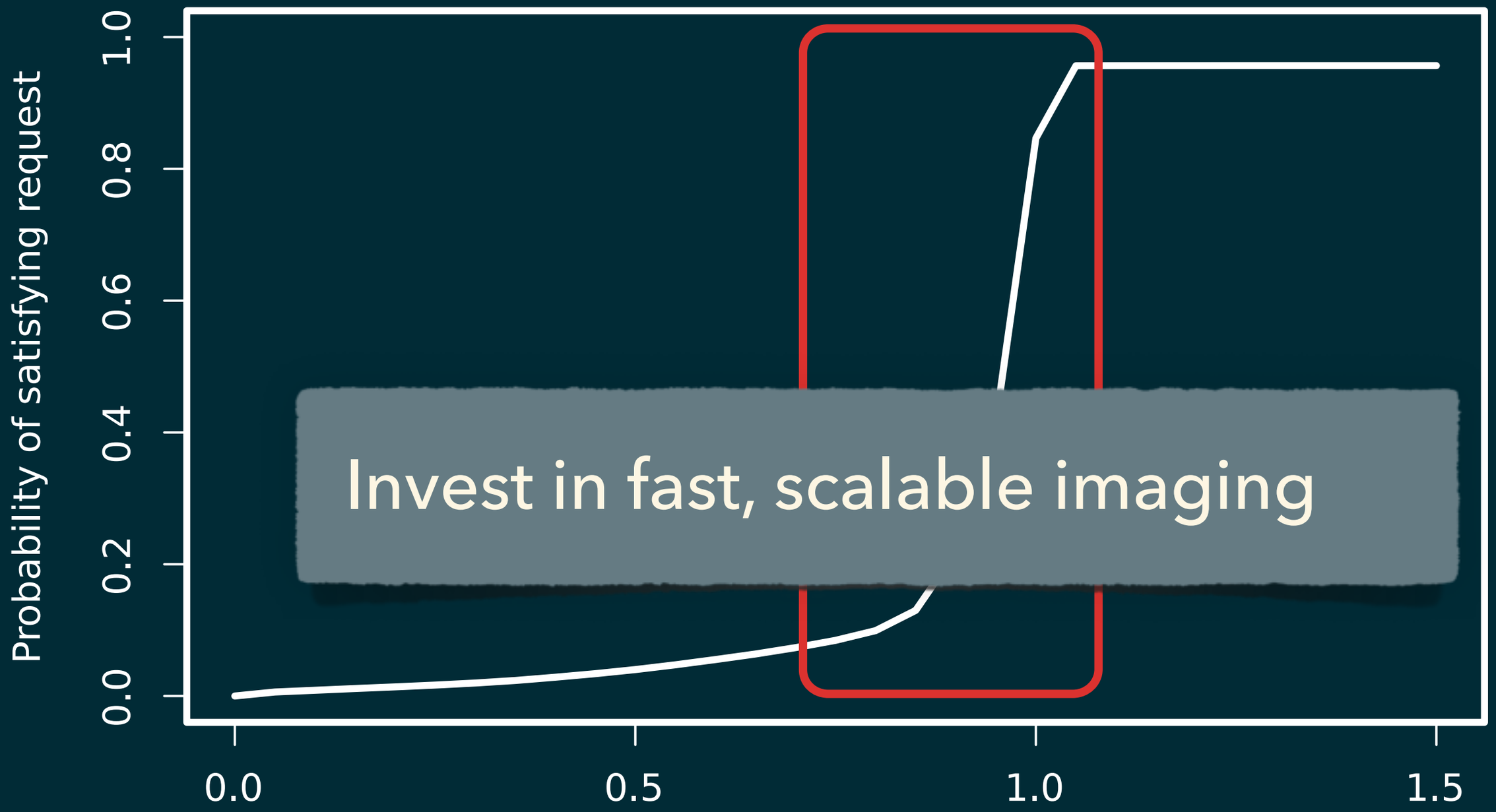
Pre-loading: Rate



Pre-loading: Rate



Pre-loading: Rate



Conclusions

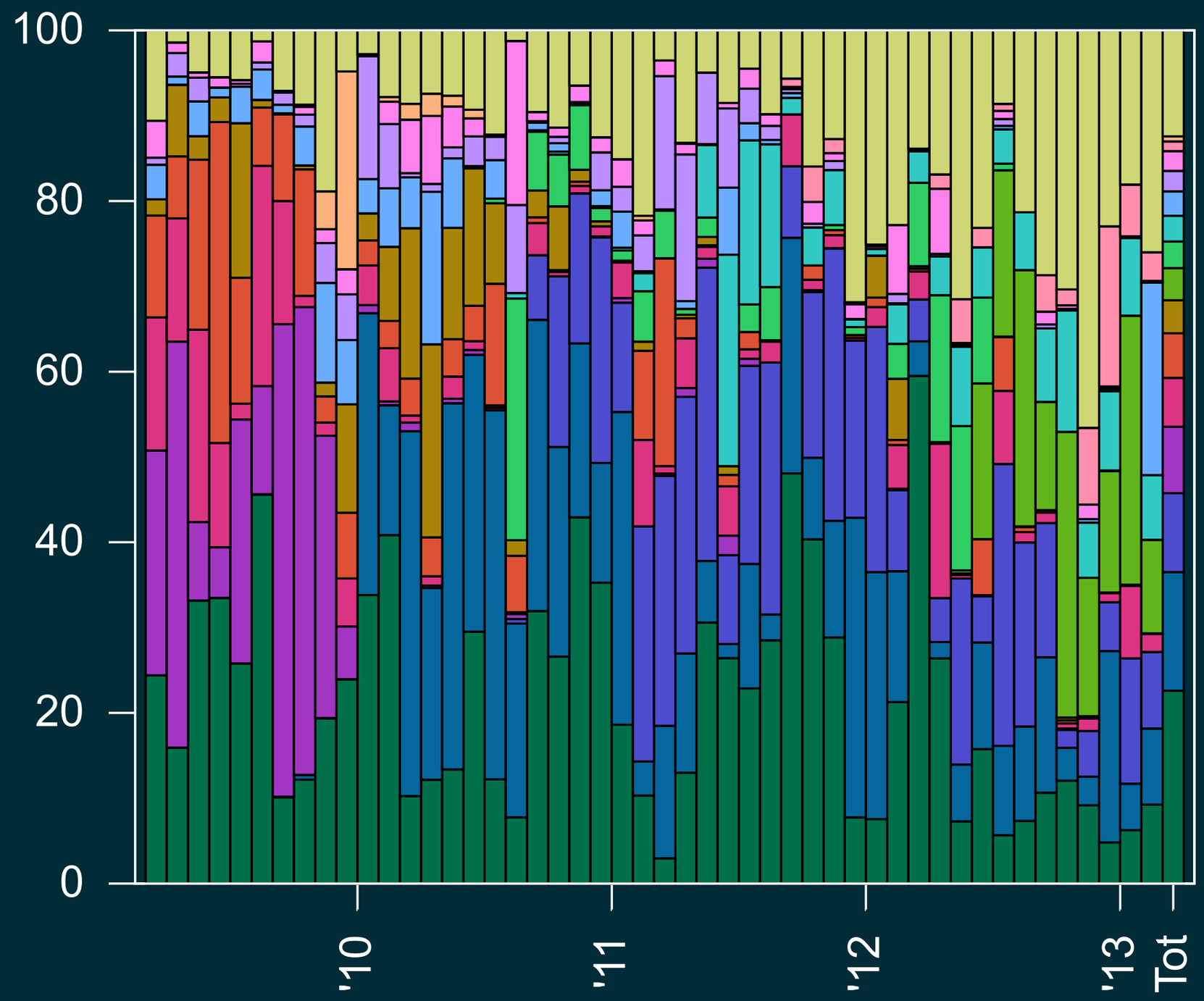
General conclusions

- Deduplicating, two-tier storage attractive
  - Caching can be effective
  - *Image lifespan, idle periods*
- Treat facility and user images differently
  - Facility better targets for pre-loading
- Differential loading requires new strategies
  - *Potential savings, outline of optimization problem*
- *Images per organization, WSS per week*

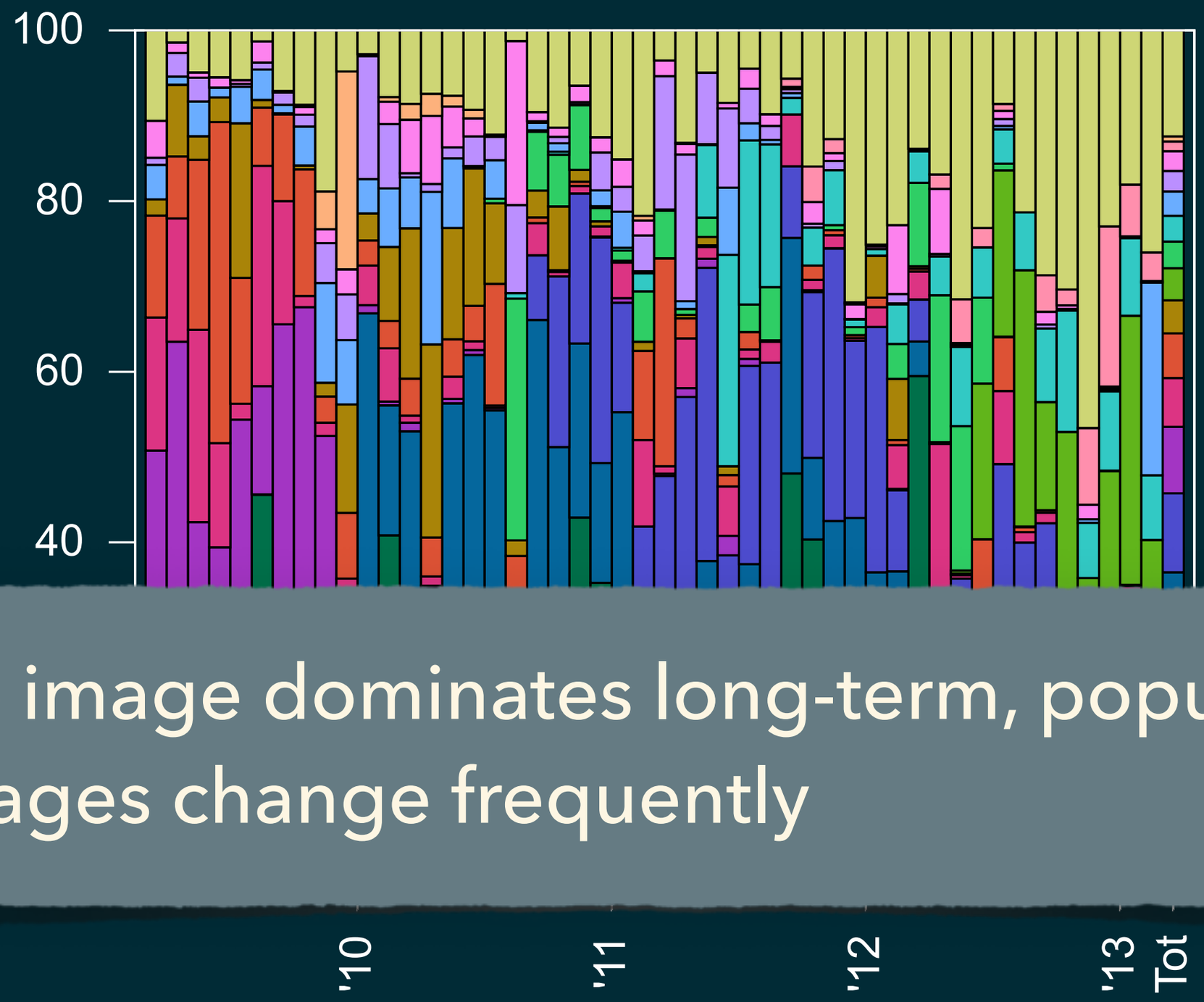
Explore the data, reproduce our results:

http://aptlab.net/p/tbres/nsdi14

No dominant images



No dominant images



No image dominates long-term, popular images change frequently

Image lifespan

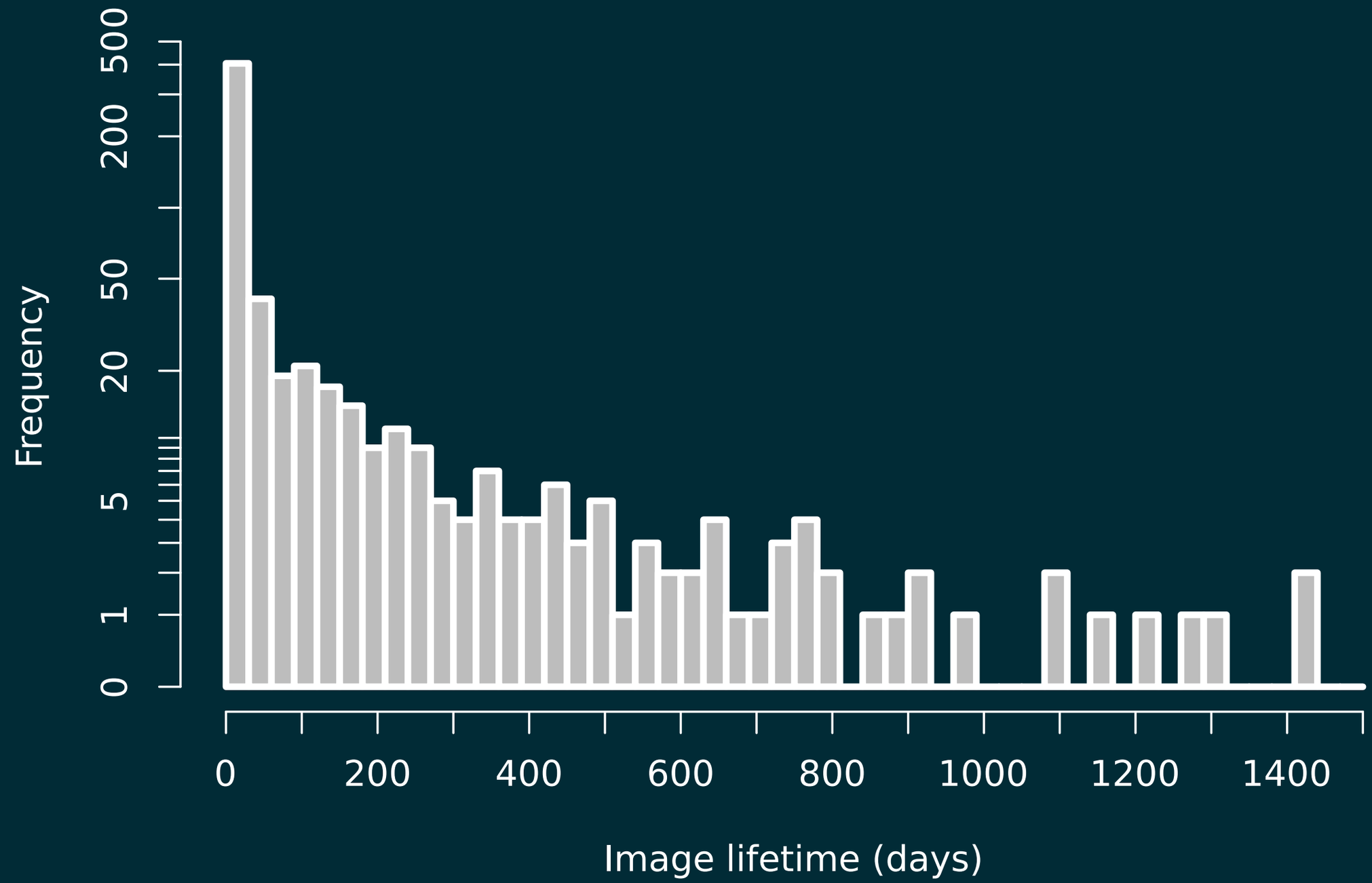


Image lifespan

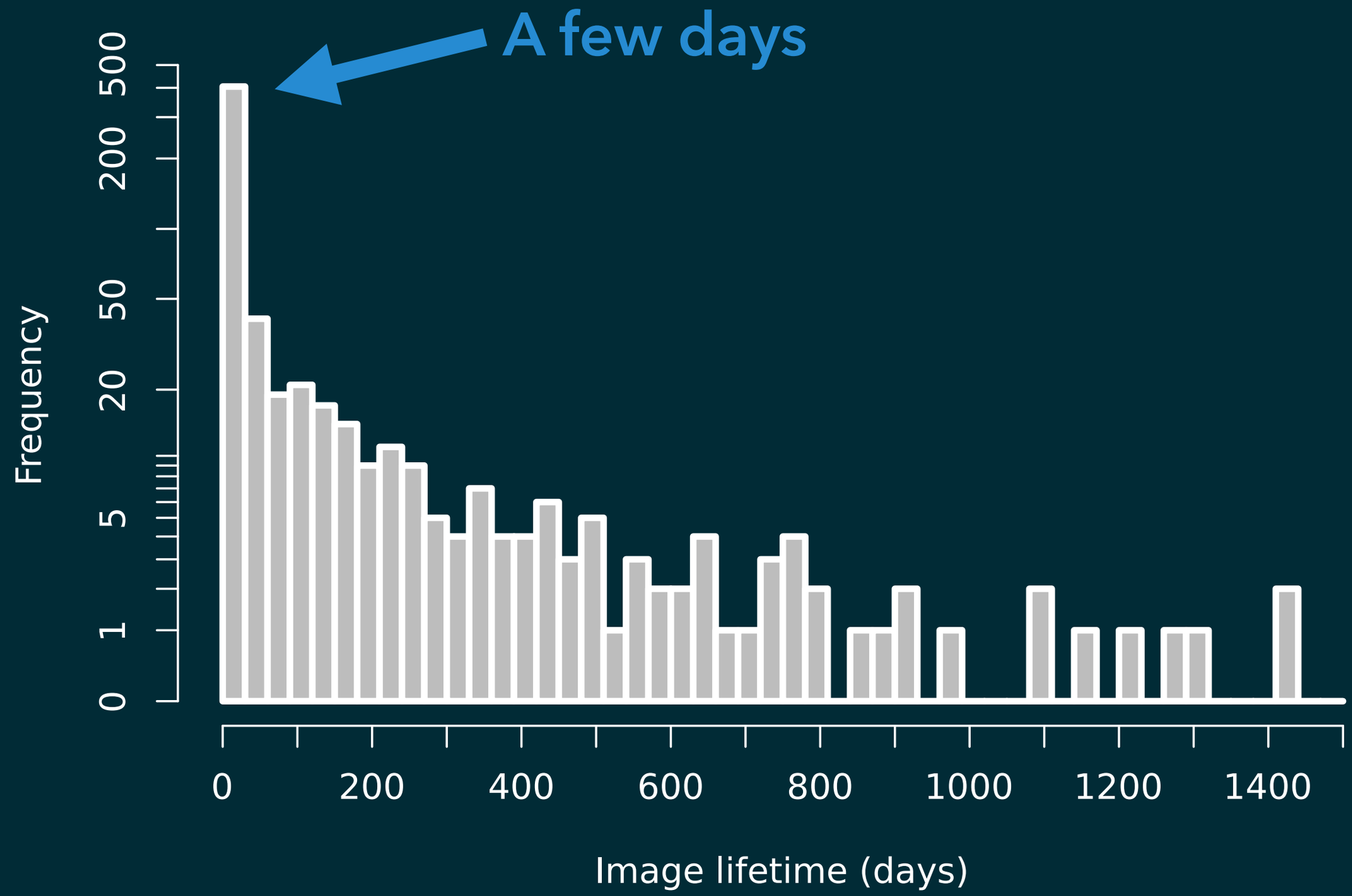


Image lifespan

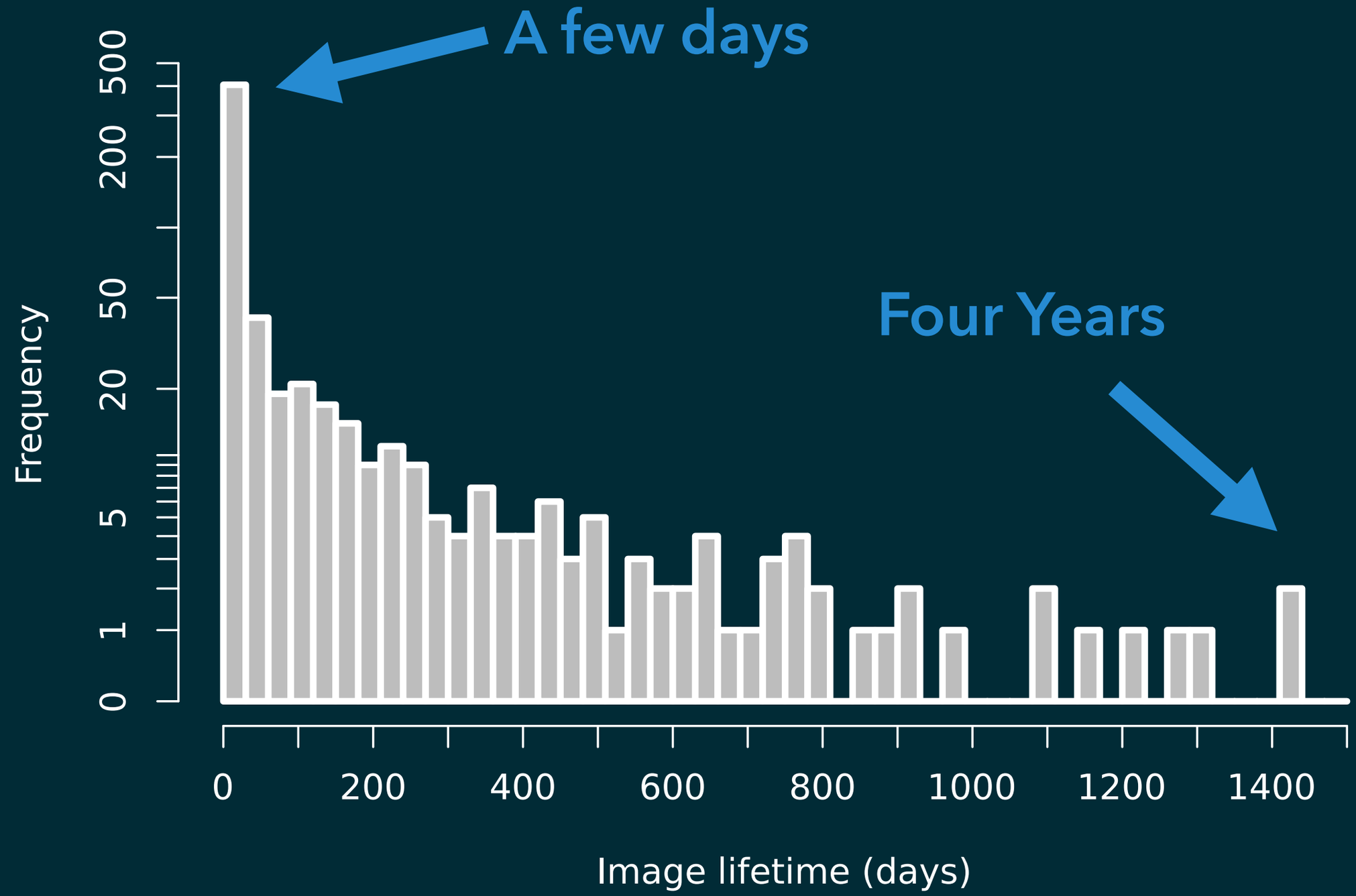
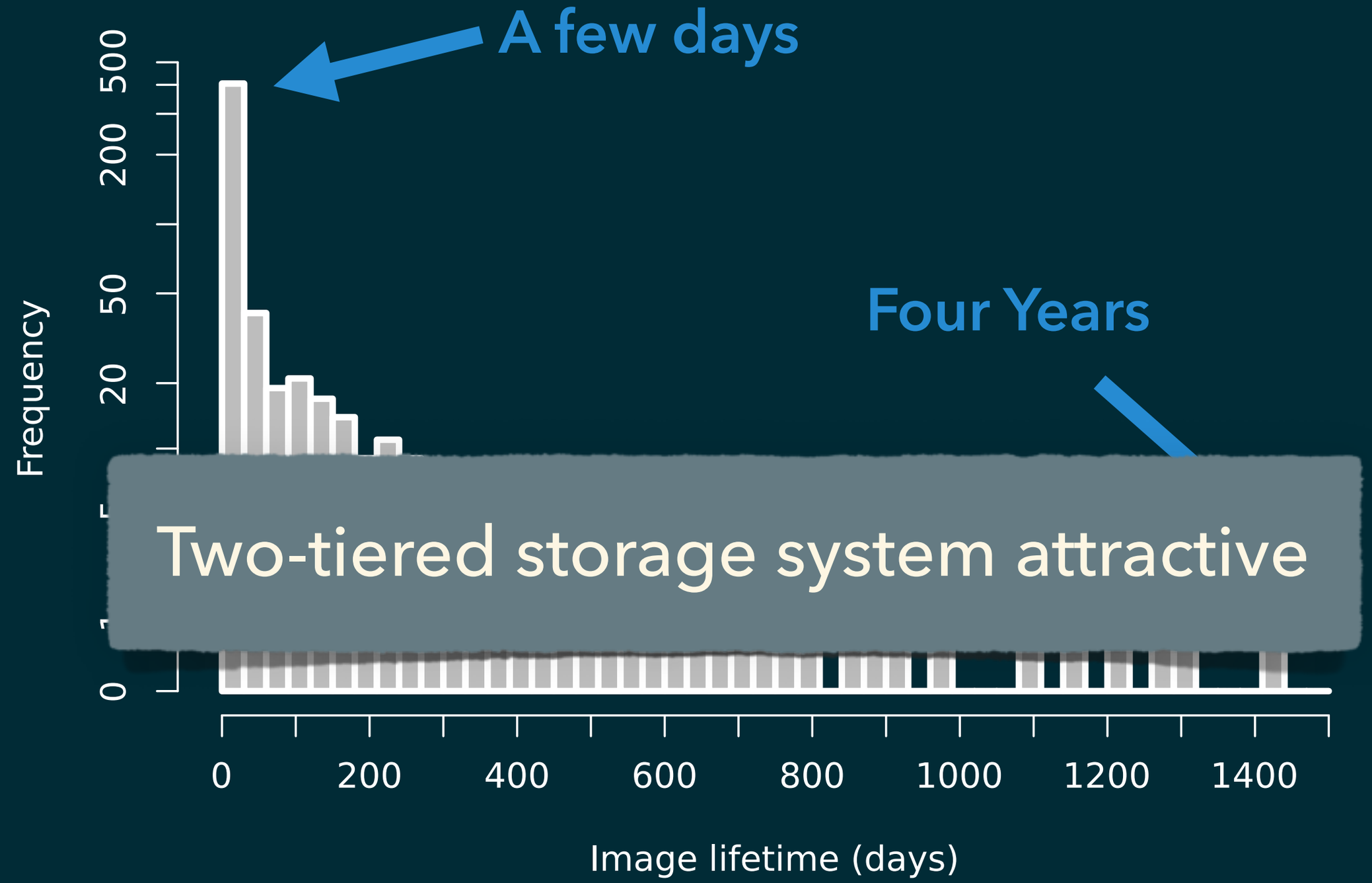
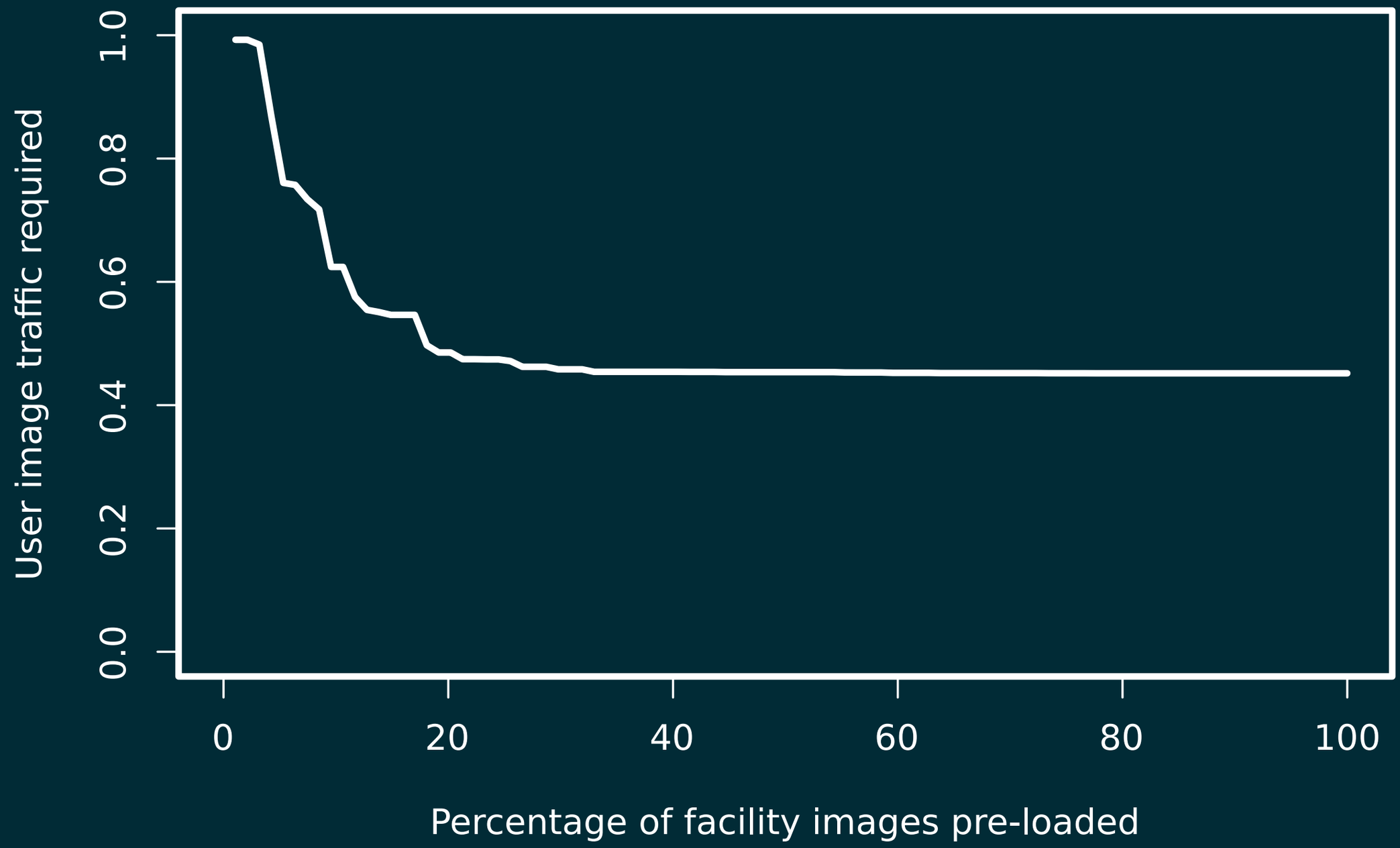


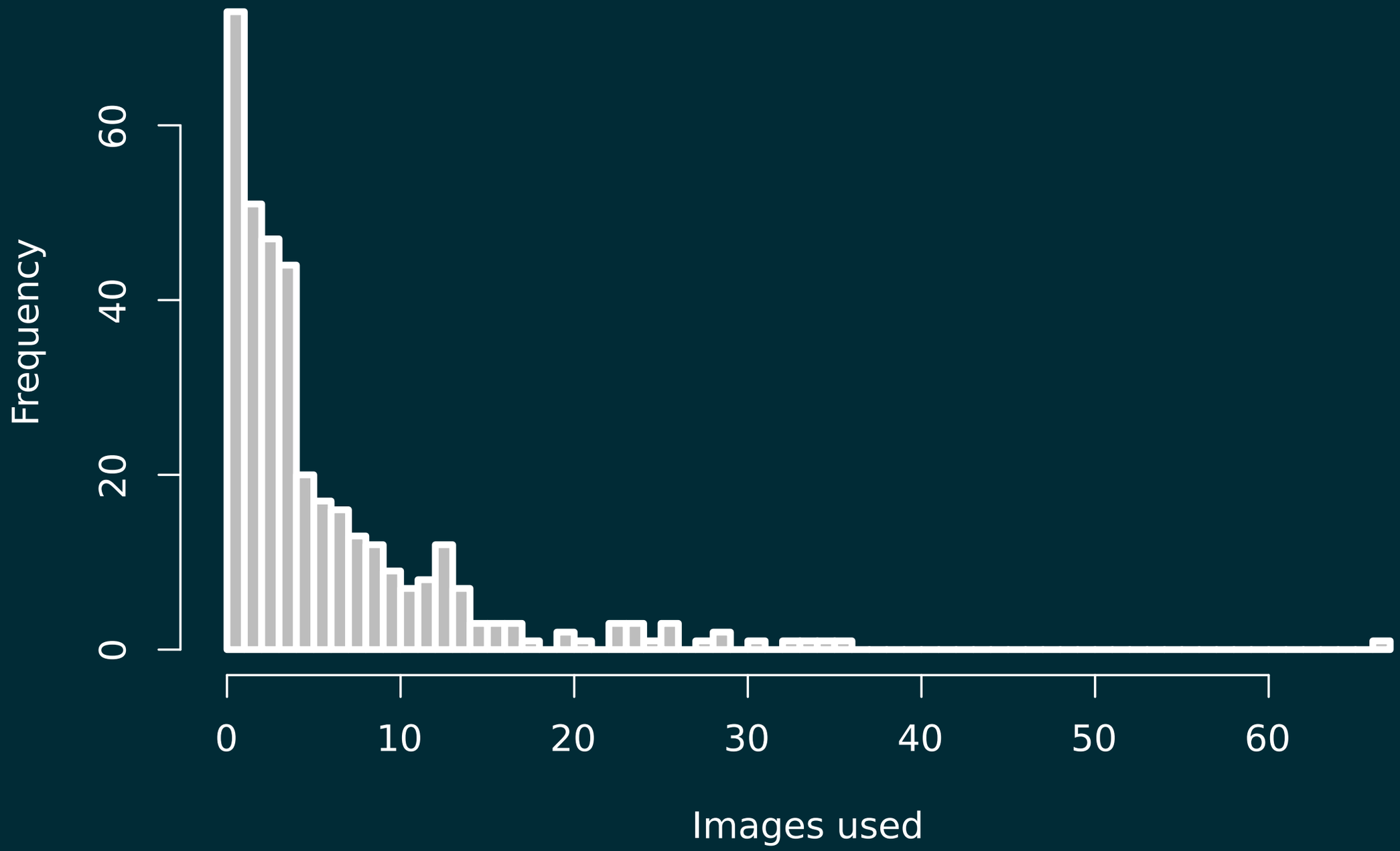
Image lifespan



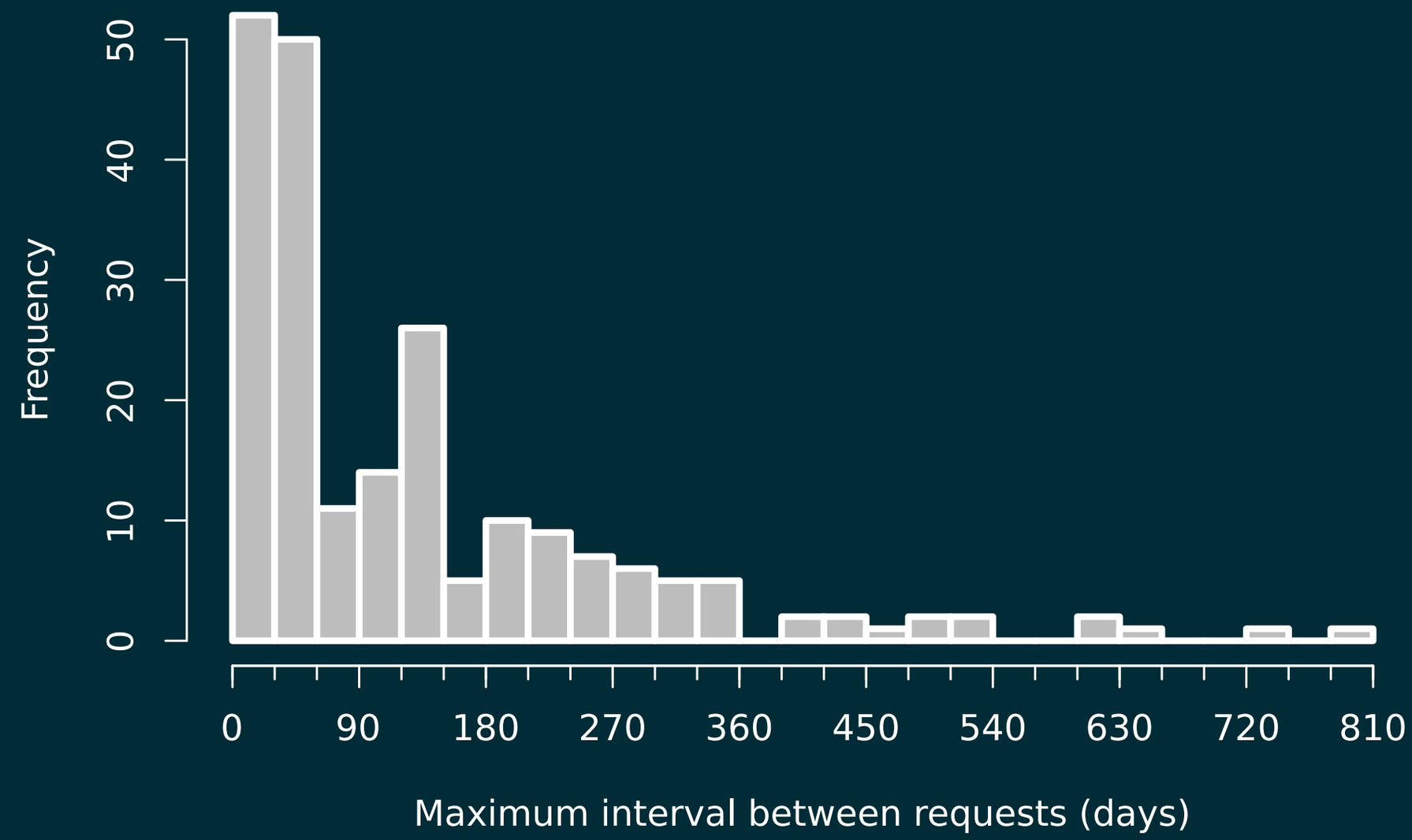
Savings from deltas



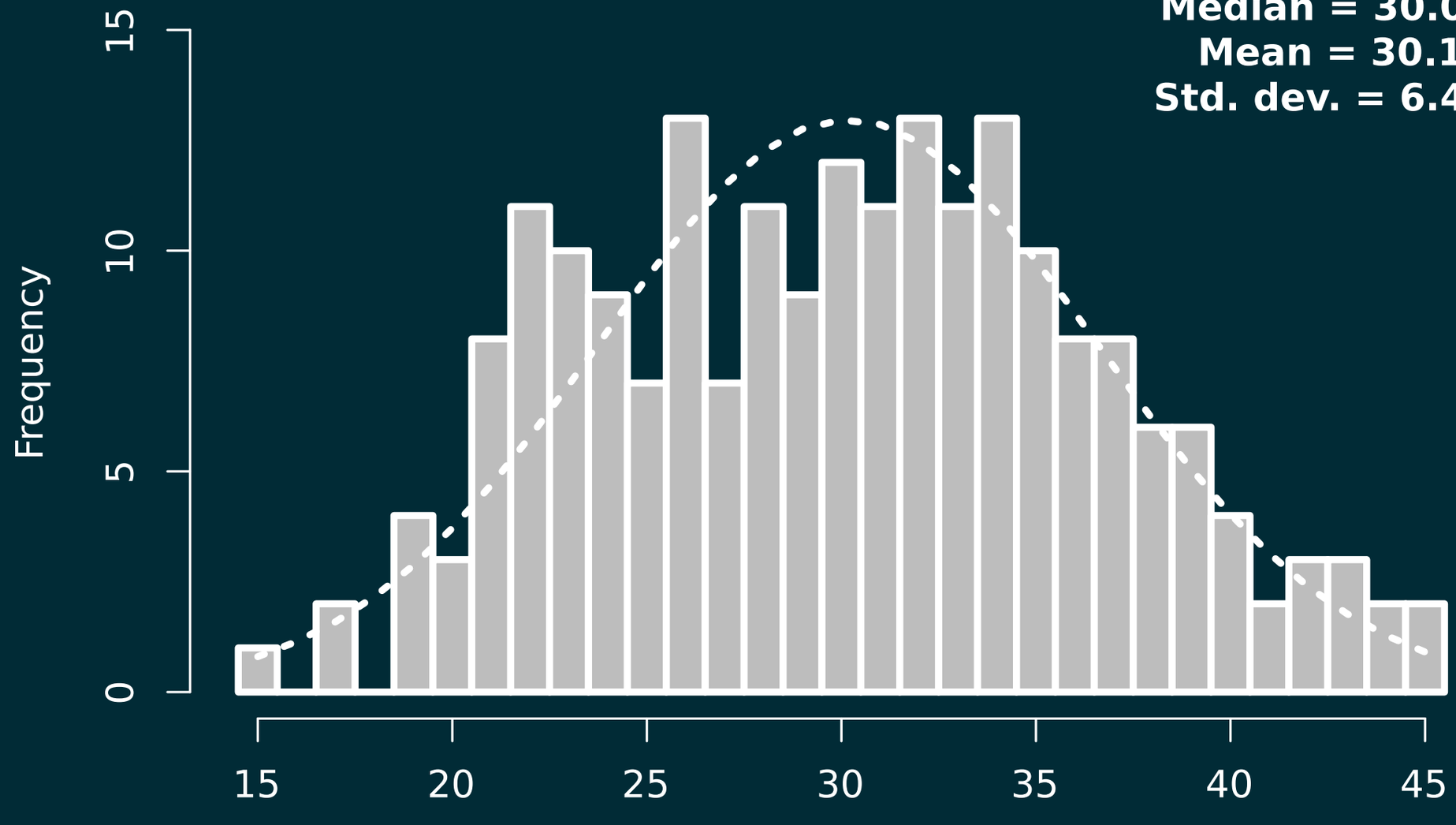
Images per organization



Idle images


Frequency
0
10
20
30
40
50
0
90
180
270
360
450
540
630
720
810
Maximum interval between requests (days)

WSS per week


Median = 30.00
Mean = 30.17
Std. dev. = 6.43
Frequency
0
5
10
15
15
20
25
30
35
40
45
Images used during week

Top images

	RHL90-STD [D]	21,993	7.9%
	FEDORA10-STD	18,042	6.4%
	UBUNTU10-STD	14,402	5.1%
	RHL90-STD	13,182	4.7%
	FC4-UPDATE	12,097	4.3%
u	715/10	11,156	4.0%
	FBSD410-STD	8,916	3.2%
	FEDORA8-STD	8,153	2.9%
u	237/69	7,512	2.7%
u	296/35	7,179	2.6%
u	787/24	6,243	2.2%
	UBUNTU70-STD	6,021	2.2%
	UBUNTU12-64-STD	5,834	2.1%

Size considerations

- Small facilities with few idle disks
  - Pre-loading not valuable
- Large facilities - focus on:
  - Scalable reloading mechanisms
  - Prediction and optimization for user requests