Steam Global Download Infrastructure—Geographic, Speed, and Temporal Patterns in a Global Gaming Platform

Author: Mingyang Wang
Date: 12/10/2025

Main Figure

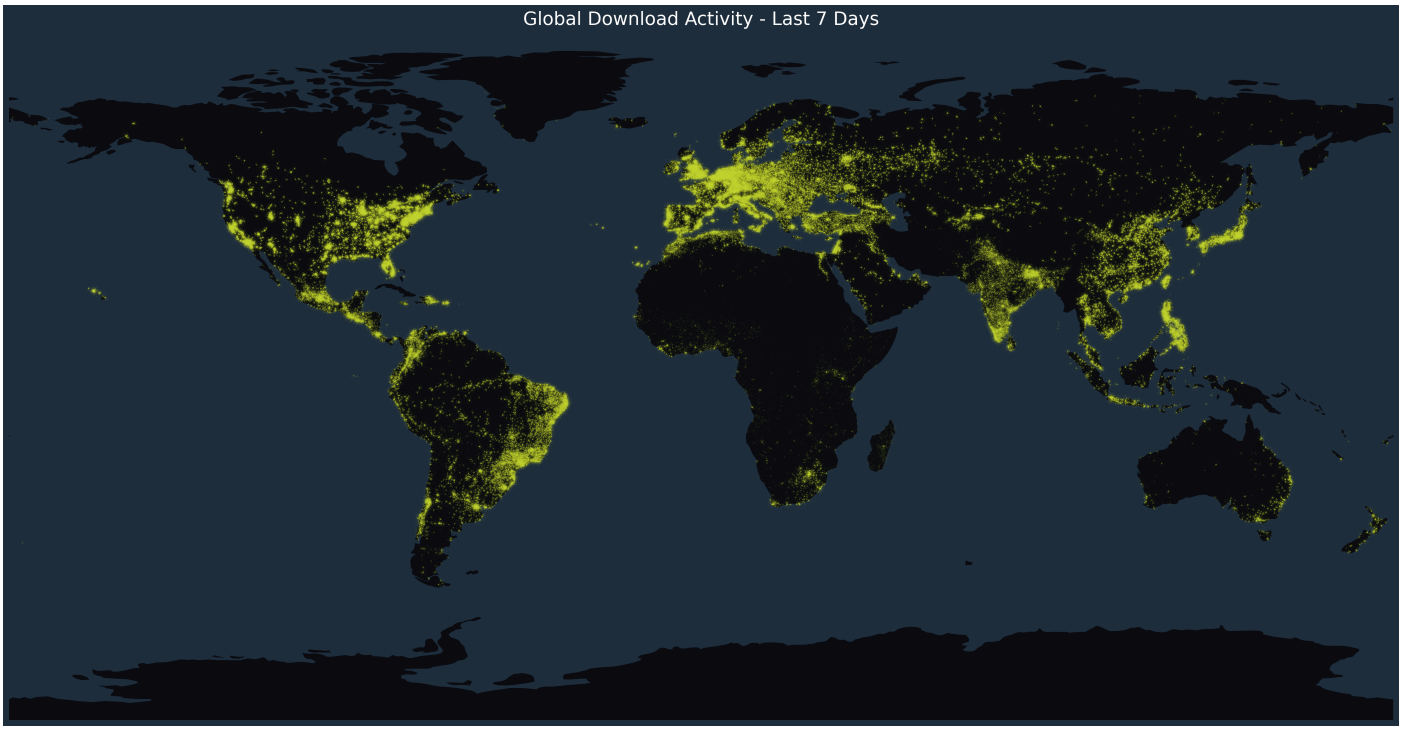

Figure 1 City-level Global Download Activity Heatmap.

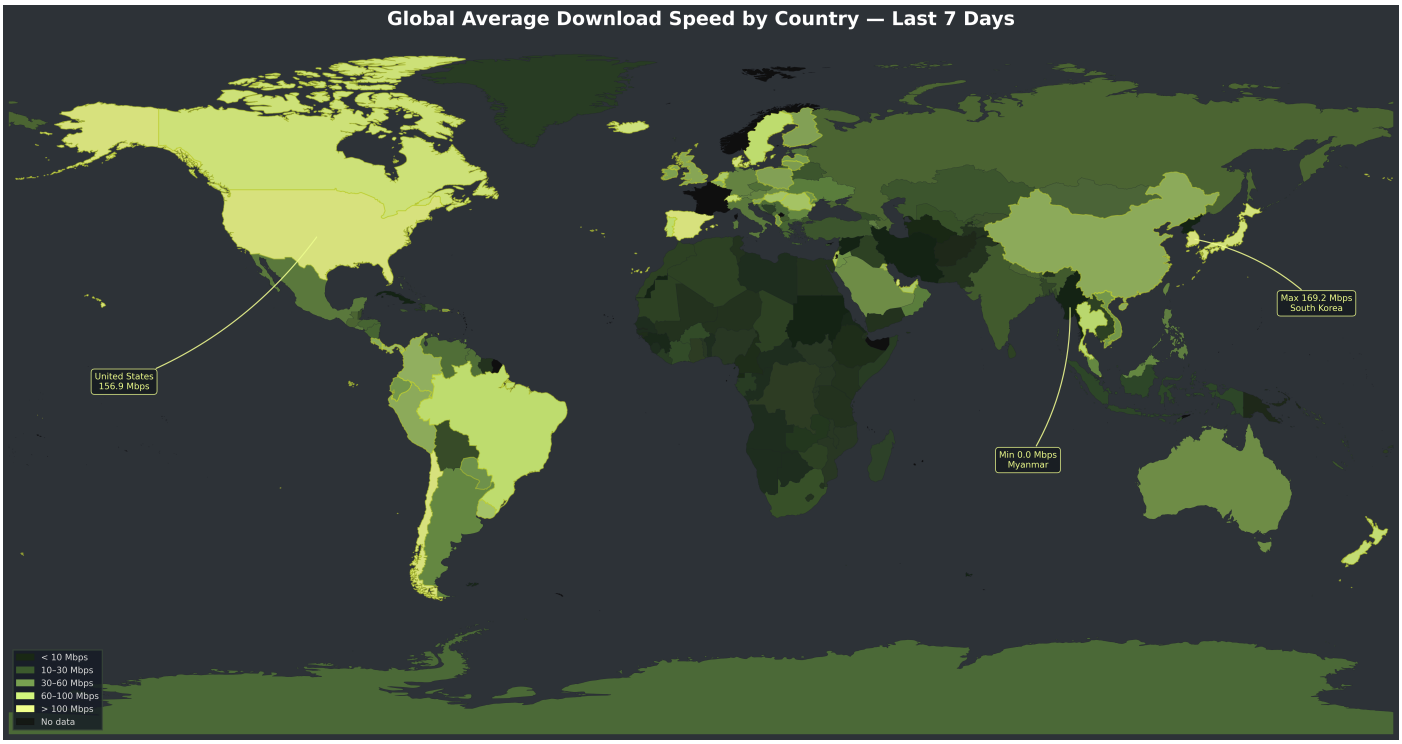

Figure 2 Average country-level download speed (Mbps) shown as a choropleth map

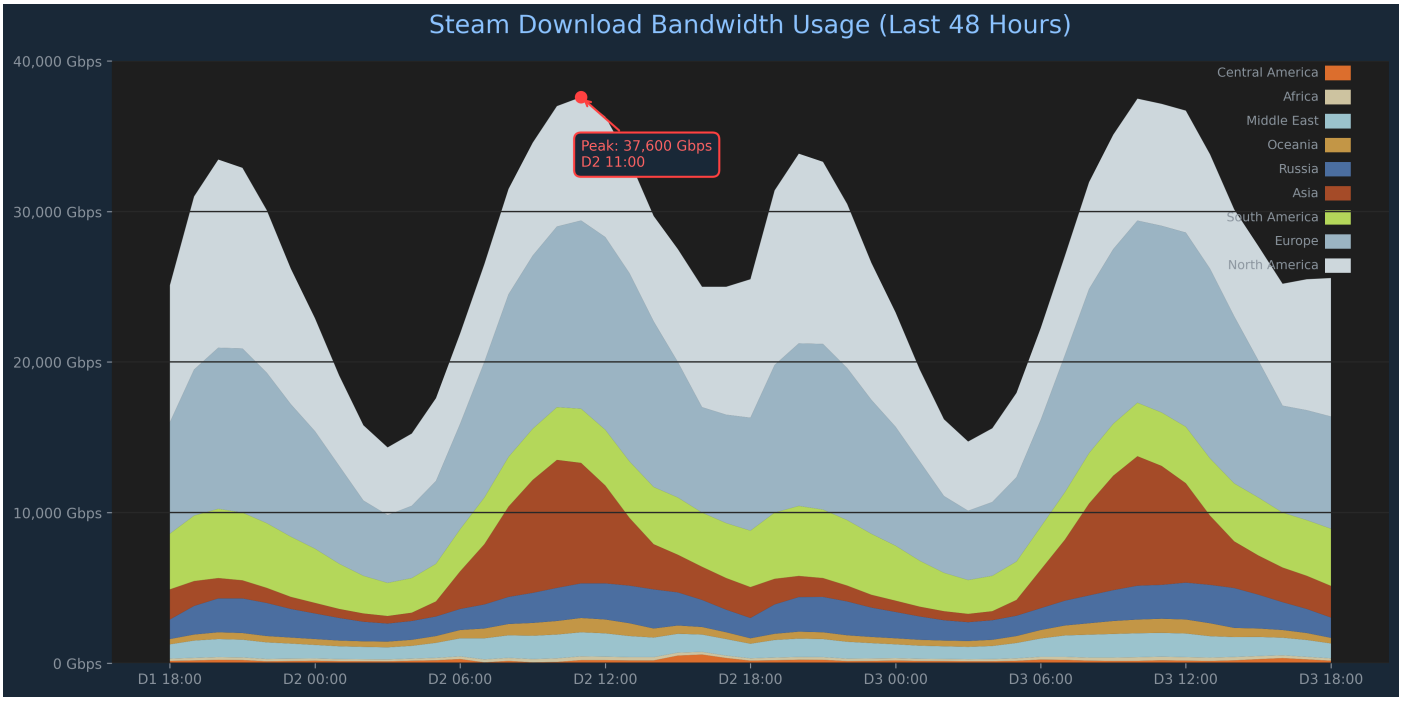

Figure 3 Stacked regional bandwidth usage over 48 hours (Gbps), revealing global diurnal demand patterns.

Figure Legend

- **Panel 1:** Each point represents population-weighted download activity allocated to cities. Higher density indicates higher activity. Bright clusters emerge from overlapping semi-transparent points.
- **Panel 2:** Country-level average download speed (Mbps). Dark green indicates slow speeds; yellow-green indicates high speeds; black indicates missing data.
- **Panel 3:** Stacked area chart of bandwidth usage by global region (Gbps). The peak point marks maximum total bandwidth demand.

Key Findings

- **Strong geographic concentration:** Download activity is highly concentrated in North America, Western Europe, and East Asia, with major urban centers forming dominant hotspots.
- **Large global speed disparities:** Average download speeds range from over 160 Mbps in South Korea and the U.S. to near-zero values in infrastructure-limited regions.
- **Clear diurnal demand cycle:** Global bandwidth usage varies by a factor of ~2.7× over 48 hours, with peaks occurring when Asia, Europe, and the Americas overlap in active hours.
- **Urban–rural divide:** Even within high-speed countries, activity clusters around major cities, revealing persistent sub-national inequality.
- **Infrastructure–population mismatch:** Large populations do not guarantee high download performance, highlighting the role of network investment rather than user count alone.

Data and Methods

Data Sources

This project uses four cleaned and standardized datasets, all stored in CSV format and matched using ISO3 country codes.
Original data came from Steam public statistics, Natural Earth, and the World Cities Database. Several files contained inconsistent or missing ISO3 codes and non-uniform units, which were manually corrected during preprocessing.

- **cities_iso3.csv**
  City-level population and geographic coordinates (>40,000 cities).
  Used to allocate country-level traffic to specific locations.

- **traffic_iso3.csv**
  Country-level Steam download traffic for the past 7 days.
  All traffic values were normalized to **TB** for consistency.
- **speed_by_country_iso3.csv**
  Average download speed (Mbps) by country from Steam's public telemetry.
  Used to generate the global speed choropleth.
- **download_bandwidth.csv**
  Hourly bandwidth usage by nine global regions over 48 hours.
  Derived from Steam CDN monitoring logs.

Methods

- **Population-weighted allocation:**
  Country-level traffic was distributed to individual cities proportionally to their population size, enabling city-level visualization of global download activity.
- **Log transformation and three-layer point cloud:**
  To prevent domination by a few high-traffic countries, values were log-transformed.
  A three-layer jitter structure (core, mid, outer) created realistic density gradients.
  *The original point cloud contained several hundred thousand points; for rendering efficiency and visual clarity, the final visualization was downsampled to approximately **300,000 points** while preserving the overall spatial distribution.*
- **Choropleth speed mapping:**
  Country-level average download speeds were mapped using a custom gradient from dark green (slow) to yellow-green (fast), with black representing missing data.
- **Temporal bandwidth visualization:**
  Regional bandwidth patterns across 48 hours were plotted using stacked area charts to highlight global demand cycles.

All figures were generated entirely in Python using `pandas`, `numpy`, `geopandas`, and `matplotlib`.

Significance Statement

Steam's global footprint offers a unique opportunity to study real-world digital infrastructure inequality. The integrated figure highlights significant disparities in bandwidth availability, network performance, and user accessibility—patterns that mirror broader inequalities affecting education, remote work, and economic participation. By combining spatial, temporal, and performance dimensions, this project demonstrates how visualization can make hidden infrastructure gaps visible and interpretable at a global scale.

Reproducibility

All Python code, datasets, and figure generation scripts are available on GitHub:
https://github.com/noob841/steam-global-download-analysis/tree/main