

NEB old vs new CSV comparison report

Generated January 29, 2026

This report compares each file prefixed with **old_** (historical extract through ~2020) to its newer version (same structure plus additional years). For overlapping rows (same keys and year), we quantify similarity and revisions using correlation and absolute percent differences.

Method

- Website-style files (wide years) are reshaped to long format using keys: region, demographic-type, demographic, year.
- Download file is compared on keys: fips, name, type, category, year (with multiple metrics).
- Overlap is defined by matching keys (including year). Statistics are computed only where both old and new values are non-missing.
- Mean/median/max abs % diff are based on $|new-old| / |old|$ (rows where old=0 can yield undefined % diff and are excluded).

Highlights

- **neb_download**: new adds 208 additional key-years; overlap keys 884. Biggest revision metric: actualization (mean abs % diff 2.98%, corr 0.9945).
- **neb_website_actualization**: new adds 208 additional key-years; overlap keys 884. Biggest revision metric: actualization (mean abs % diff 2.98%, corr 0.9945).
- **neb_website_bf_per_capita**: new adds 52 additional key-years; overlap keys 832. Biggest revision metric: bf_per_capita (mean abs % diff 2.70%, corr 0.9901).
- **neb_website_index**: not a pure append (keys only old 13, only new 135); overlap keys 585. Biggest revision metric: index (mean abs % diff 0.37%, corr 0.9996).
- **neb_website_newness**: new adds 208 additional key-years; overlap keys 728. Biggest revision metric: newness (mean abs % diff 2.01%, corr 0.9971).

Coverage and similarity tables

Coverage is computed on unique keys (including year). Similarity metrics are computed on overlapping rows with non-missing values.

Table 1. Coverage / overlap by dataset

Dataset pair	Old rows	New rows	Overlap rows	Unique keys (old)	Unique keys (new)	Keys in both	Keys only old	Keys only new
neb_download	884	1092	884	884	1092	884	0	208
neb_website_actualization	884	1092	884	884	1092	884	0	208
neb_website_bf_per_capita	832	884	832	832	884	832	0	52
neb_website_index	598	720	585	598	720	585	13	135
neb_website_newness	728	936	728	728	936	728	0	208

Table 2. Similarity on overlapping rows (old vs new)

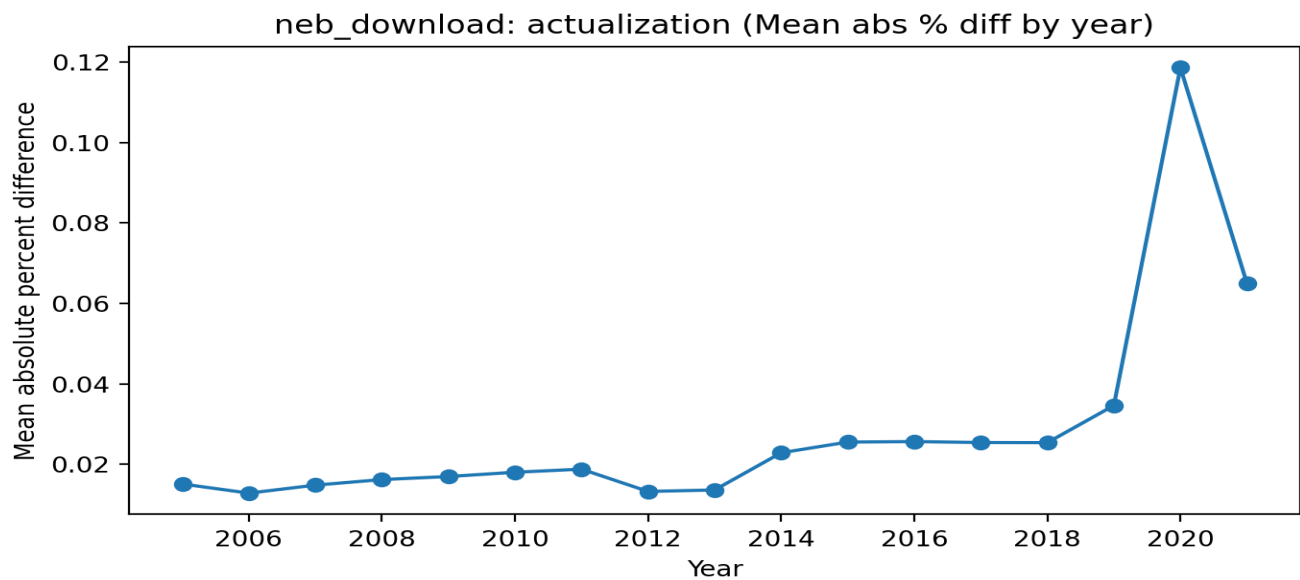
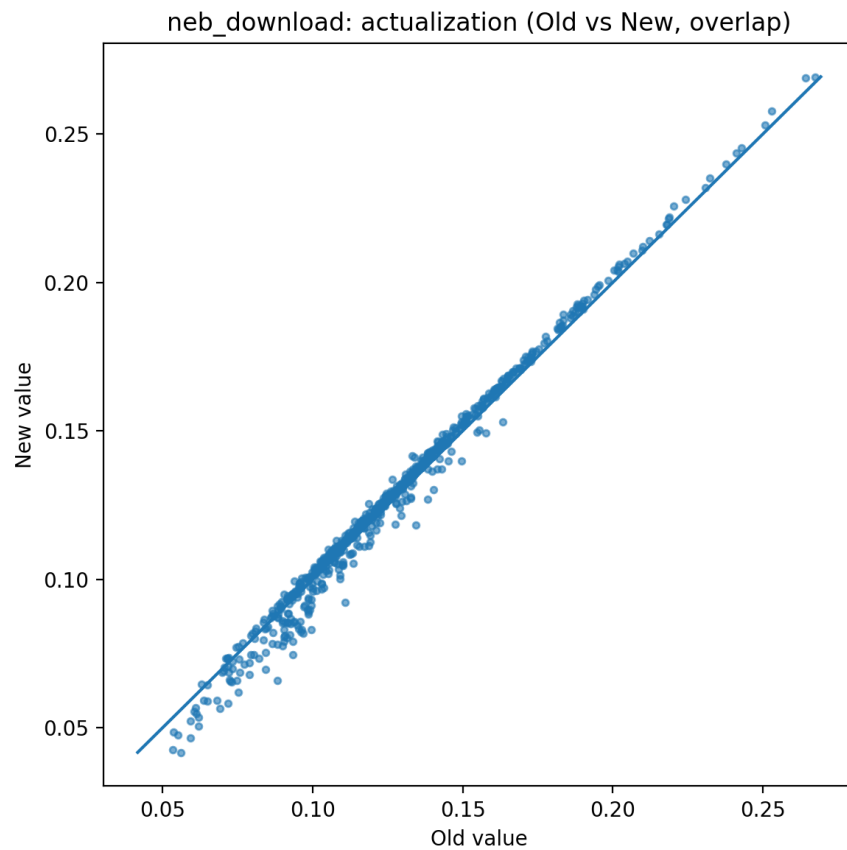
Dataset pair	Metric	Overlap rows	Corr (r)	Mean abs % diff	Median abs % diff	Max abs % diff	Share >5%	Share >10%
neb_download	actualization	692	0.9945	2.98%	1.96%	25.56%	13.01%	5.49%
neb_download	bf_per_capita	640	0.9901	2.70%	1.89%	26.12%	9.06%	4.53%
neb_download	index	512	0.9996	0.37%	0.29%	1.85%	0.00%	0.00%
neb_download	newness	551	0.9971	2.01%	1.80%	9.25%	3.09%	0.00%
neb_download	velocity	512	0.9996	0.90%	0.86%	2.49%	0.00%	0.00%
neb_website_actualization	actualization	692	0.9945	2.98%	1.96%	25.56%	13.01%	5.49%
neb_website_bf_per_capita	bf_per_capita	640	0.9901	2.70%	1.89%	26.12%	9.06%	4.53%
neb_website_index	index	512	0.9996	0.37%	0.29%	1.85%	0.00%	0.00%
neb_website_newness	newness	551	0.9971	2.01%	1.80%	9.25%	3.09%	0.00%

Note: Overlap rows can differ by metric because some values are missing in one file but not the other.

NEB DOWNLOAD

Coverage: 884 key-years in old; 1092 in new; overlap 884. Keys only new: 208. Keys only old: 0.

Main revision signal: actualization mean abs % diff 2.98% (max 25.56%), corr 0.9945.

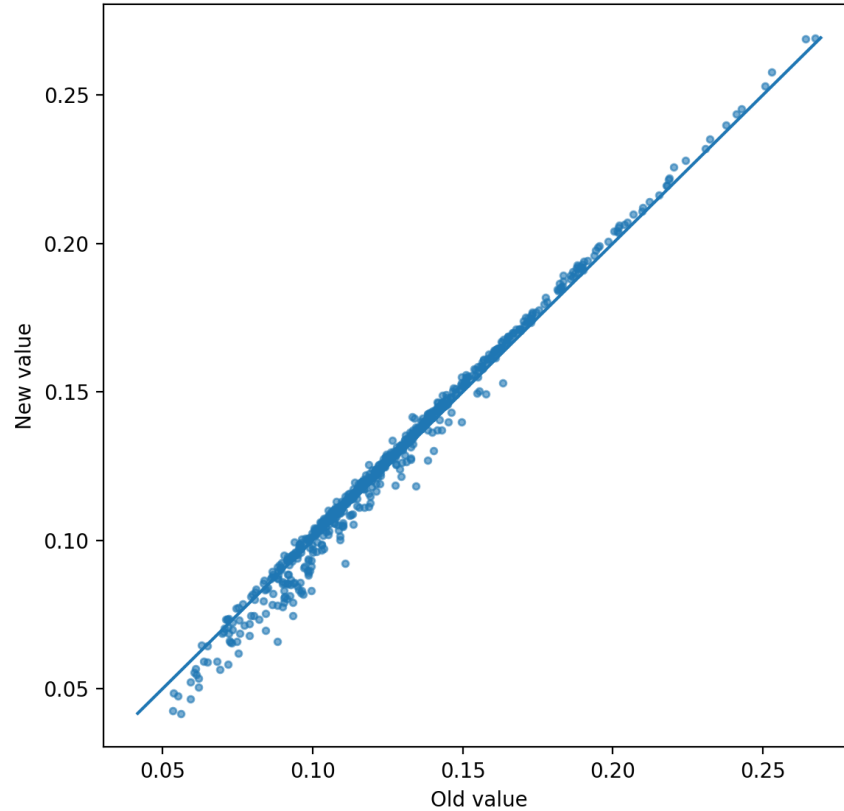


NEB WEBSITE ACTUALIZATION

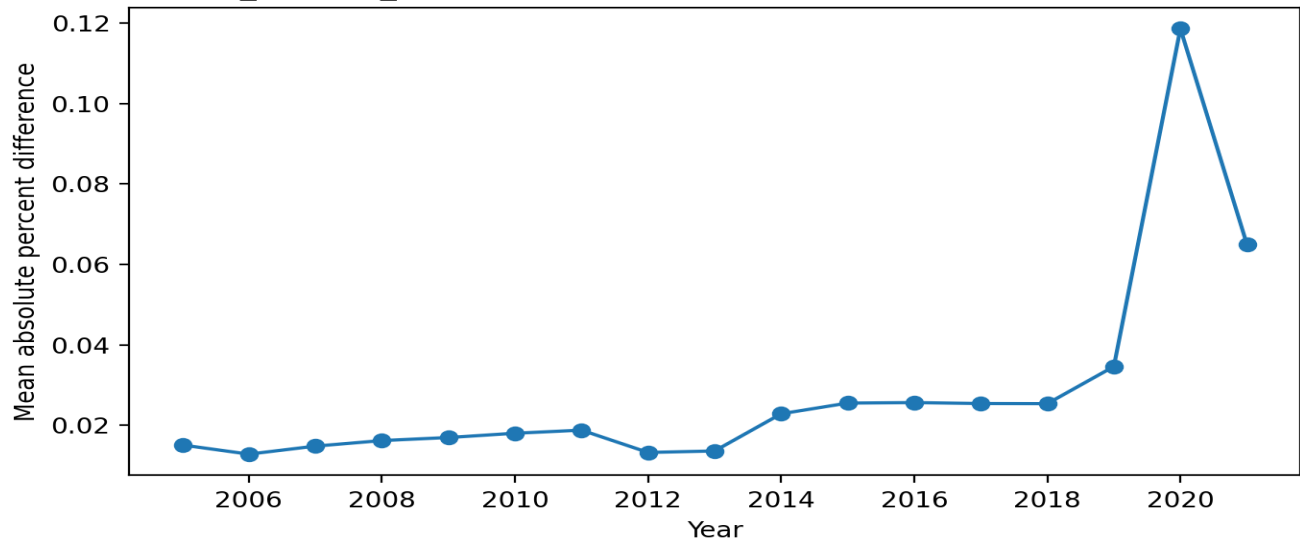
Coverage: 884 key-years in old; 1092 in new; overlap 884. Keys only new: 208. Keys only old: 0.

Main revision signal: actualization mean abs % diff 2.98% (max 25.56%), corr 0.9945.

neb_website_actualization: actualization (Old vs New, overlap)



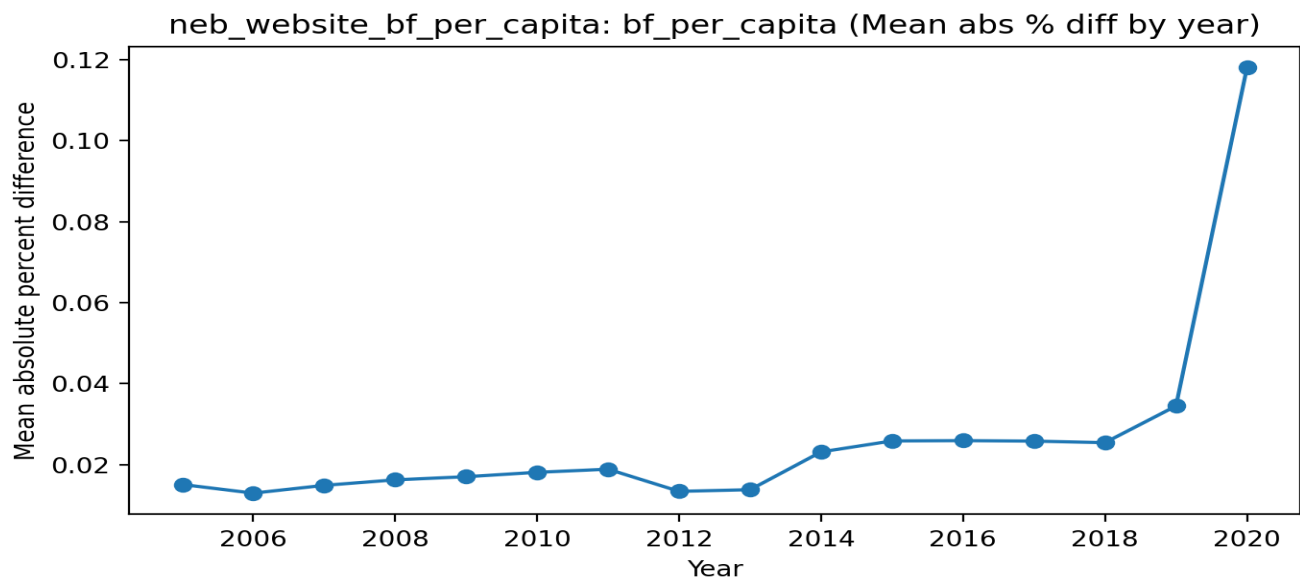
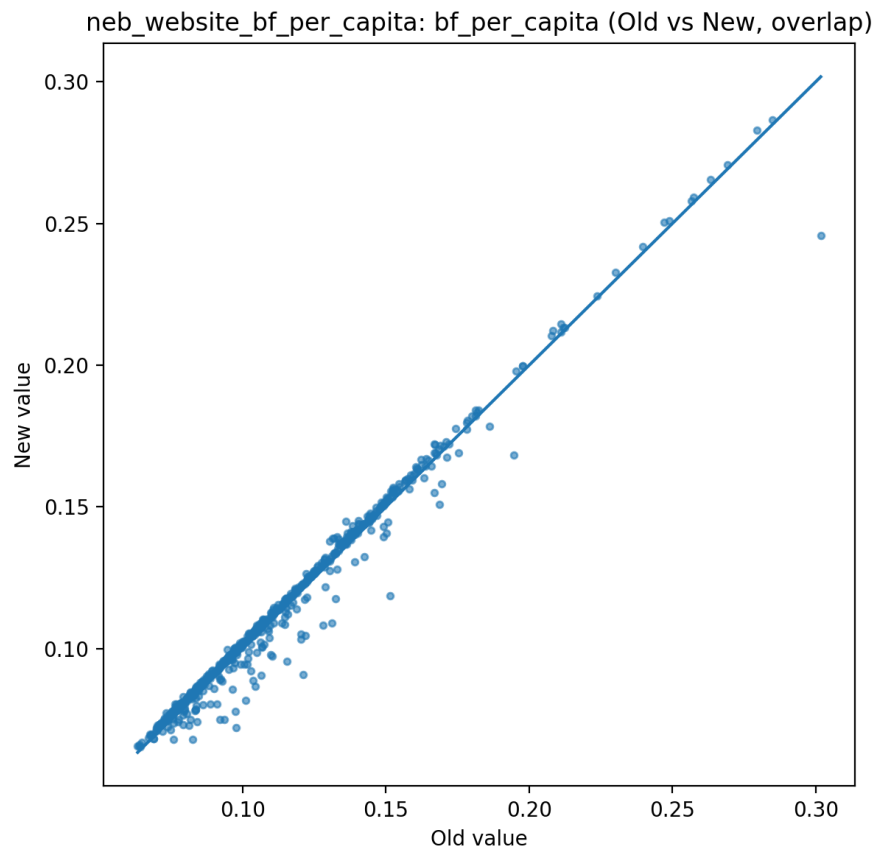
neb_website_actualization: actualization (Mean abs % diff by year)



NEB WEBSITE BF PER CAPITA

Coverage: 832 key-years in old; 884 in new; overlap 832. Keys only new: 52. Keys only old: 0.

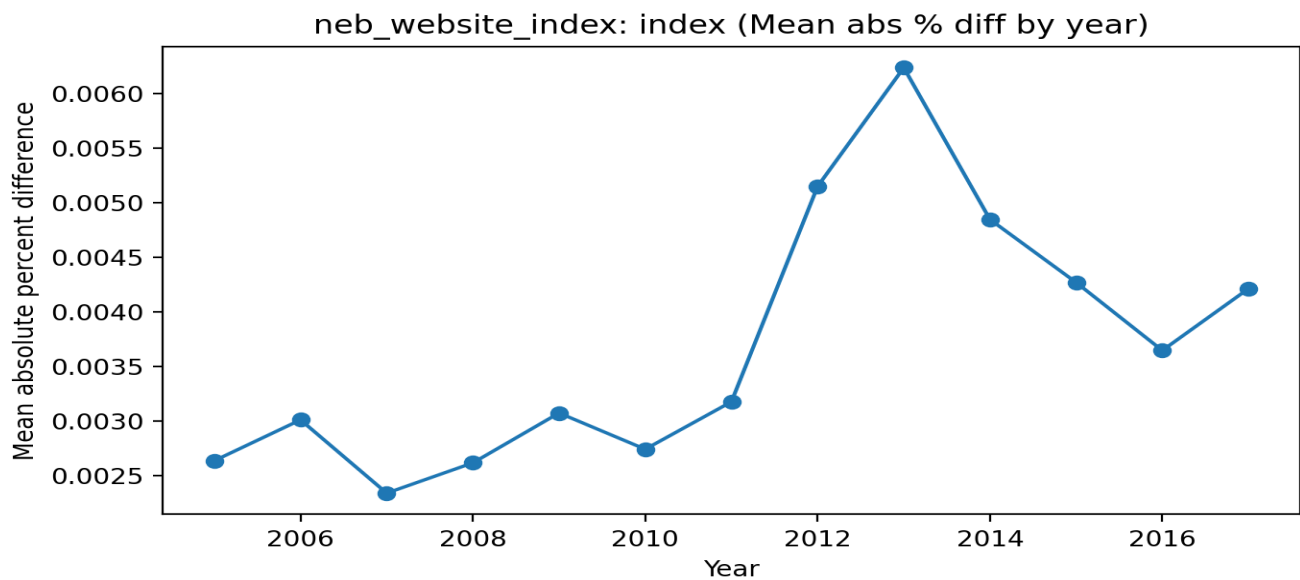
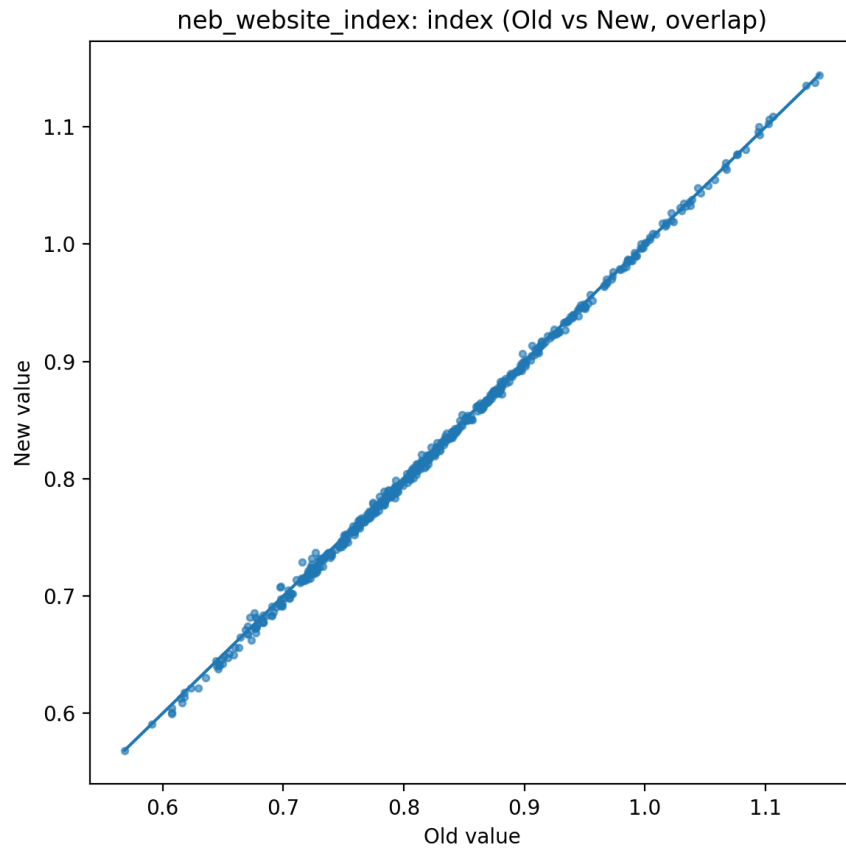
Main revision signal: bf_per_capita mean abs % diff 2.70% (max 26.12%), corr 0.9901.



NEB WEBSITE INDEX

Coverage: 598 key-years in old; 720 in new; overlap 585. Keys only new: 135. Keys only old: 13.

Main revision signal: index mean abs % diff 0.37% (max 1.85%), corr 0.9996.



NEB WEBSITE NEWNESS

Coverage: 728 key-years in old; 936 in new; overlap 728. Keys only new: 208. Keys only old: 0.

Main revision signal: newness mean abs % diff 2.01% (max 9.25%), corr 0.9971.

