

EJI - Old vs New CSV Comparison Report

Website files are compared cell-by-cell across overlapping years for matching demographic rows; download files are compared row-by-row on shared keys (including year).

Coverage summary

dataset	keys_old	keys_new	keys_both	only_old	only_new	rows_old	rows_new	overlap
eji_download	82440	98904	79900	2540	19004	412200	494520	1997500
eji_website_compensation	24732	24726	24360	372	366	24732	24726	487200
eji_website_constancy	24732	24726	24360	372	366	24732	24726	487200
eji_website_contribution	24732	24726	24360	372	366	24732	24726	487200
eji_website_creation	24732	24726	24360	372	366	24732	24726	487200

Similarity metrics (core)

dataset	metric	n	corr	MAE	RMSE	MAPE
eji_download	constancy	1911783	0.2470	0.0939	0.1268	15.61%
eji_download	compensation	1939431	0.3639	0.1550	0.2695	26.57%
eji_download	demographic-code	1997500	0.0000	1.6000	2.0000	77.00%
eji_download	contribution	1864411	0.0000	0.2498	0.4152	369.43%
eji_download	creation	418670	0.3000	3.9926	10.5127	769.33%
eji_website_compensation	compensation	473349	0.8776	0.0318	0.1187	4.63%
eji_website_constancy	constancy	467401	0.9724	0.0113	0.0253	1.97%
eji_website_contribution	contribution	455176	0.9991	0.0043	0.0116	6.15%
eji_website_creation	creation	107368	0.9899	0.4424	1.2032	70.87%

Similarity metrics (revision distribution)

dataset	metric	med_APE	max_APE	>1%	>5%	>10%
eji_download	constancy	11.75%	1653.20%	85.64%	68.95%	54.74%
eji_download	compensation	18.79%	13867.43%	90.50%	75.77%	65.23%
eji_download	demographic-code	50.00%	400.00%	80.00%	80.00%	80.00%
eji_download	contribution	63.64%	92340.11%	93.43%	83.24%	77.23%
eji_download	creation	110.60%	5651684.53%	98.13%	93.84%	90.27%
eji_website_compensation	compensation	1.59%	1466.12%	61.61%	21.26%	10.13%
eji_website_constancy	constancy	0.75%	1429.38%	42.29%	8.79%	3.23%
eji_website_contribution	contribution	2.59%	3193.88%	71.73%	31.62%	15.43%
eji_website_creation	creation	12.27%	65385.37%	91.92%	69.64%	54.50%

ejl_download

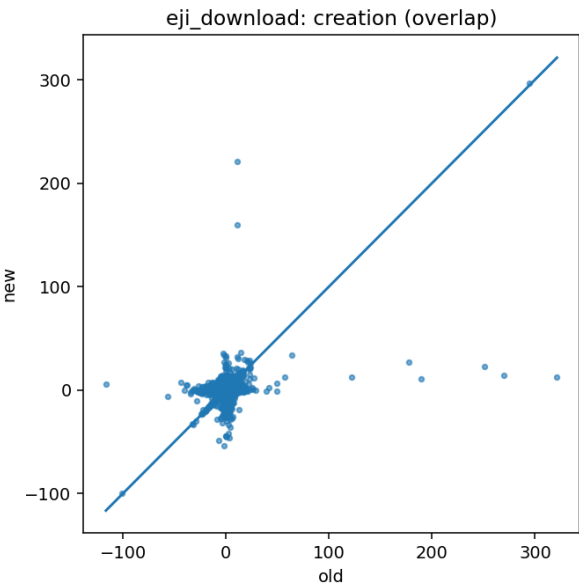
Key coverage

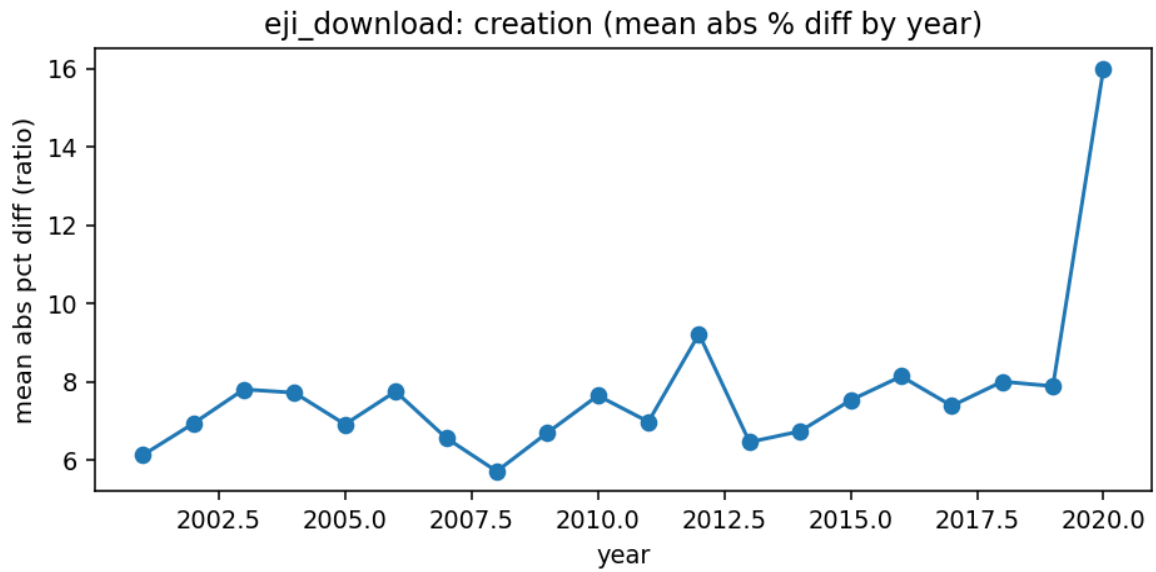
unique_keys_old	unique_keys_new	keys_in_both	keys_only_old	keys_only_new	old_rows	new_rows	overlap_rows
82440	98904	79900	2540	19004	412200	494520	1997500

Similarity stats (overlap only)

dataset_pair	metric	overlap_rows	corr	MAE	RMSE	MAPE	Median_APE	Max_APE	%>1%	%>5%	%>10%
ejl_download	creation	418670	0.3000	3.9926	10.5127	769.33%	110.60%	5651684.53%	98.13%	93.84%	90.27%
ejl_download	contribution	1864411	0.0000	0.2498	0.4152	369.43%	63.64%	92340.11%	93.43%	83.24%	77.23%
ejl_download	demographic-code	1997500	0.0000	1.6000	2.0000	77.00%	50.00%	400.00%	80.00%	80.00%	80.00%
ejl_download	compensation	1939431	0.3639	0.1550	0.2695	26.57%	18.79%	13867.43%	90.50%	75.77%	65.23%
ejl_download	constancy	1911783	0.2470	0.0939	0.1268	15.61%	11.75%	1653.20%	85.64%	68.95%	54.74%

Charts for primary metric: creation





Largest revisions (top 12 by abs % diff)

fips	name	year	creation_old	creation_new	abs_diff	abs_pct_diff
1	Alabama	2020	-0.0002031887631737	-11.483791083576026	11.4836	5651684.53%
18	Indiana	2020	-0.0002960790400762	-15.364412943559213	15.3641	5189194.37%
53	Washington	2012	-0.0001449893563313	4.619923552593795	4.6201	3186488.07%
53	Washington	2012	-0.0001449893563313	3.661757988832225	3.6619	2525635.72%
1	Alabama	2020	-0.0002031887631737	3.2681686453700247	3.2684	1608539.66%
18	Indiana	2020	-0.0002960790400762	3.557471566514204	3.5578	1201627.66%
55	Wisconsin	2010	-0.0003514645086745	4.121402932376517	4.1218	1172737.02%
41940	San Jose-Sunnyvale-Santa Clara, CA	2004	-0.0005820667560721	6.153027678438384	6.1536	1057200.00%
18140	Columbus, OH	2018	-0.0004752413513202	4.420596744975646	4.4211	930279.31%
9	Connecticut	2007	0.0005670107476887	4.642967507449104	4.6424	818750.00%
30	Montana	2020	-0.0009254315055752	-7.512669812275232	7.5117	811701.82%
30	Montana	2020	-0.0009254315055752	5.92147015810822	5.9224	639960.45%

eji_website_compensation

Key coverage

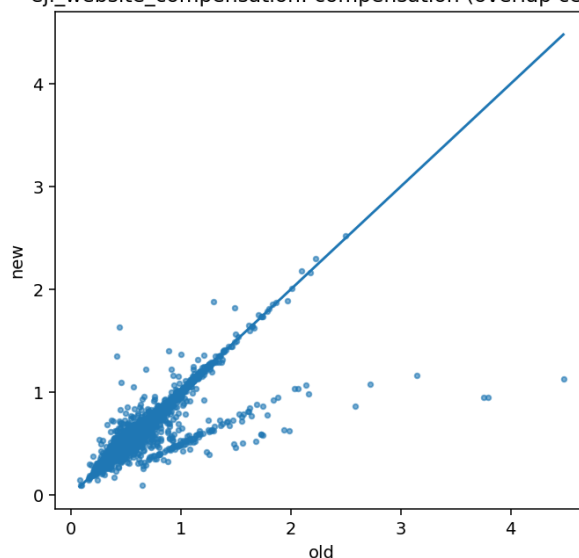
unique_keys_old	unique_keys_new	keys_in_both	keys_only_old	keys_only_new	old_rows	new_rows	overlap_rows
24732	24726	24360	372	366	24732	24726	487200

Similarity stats (overlap only)

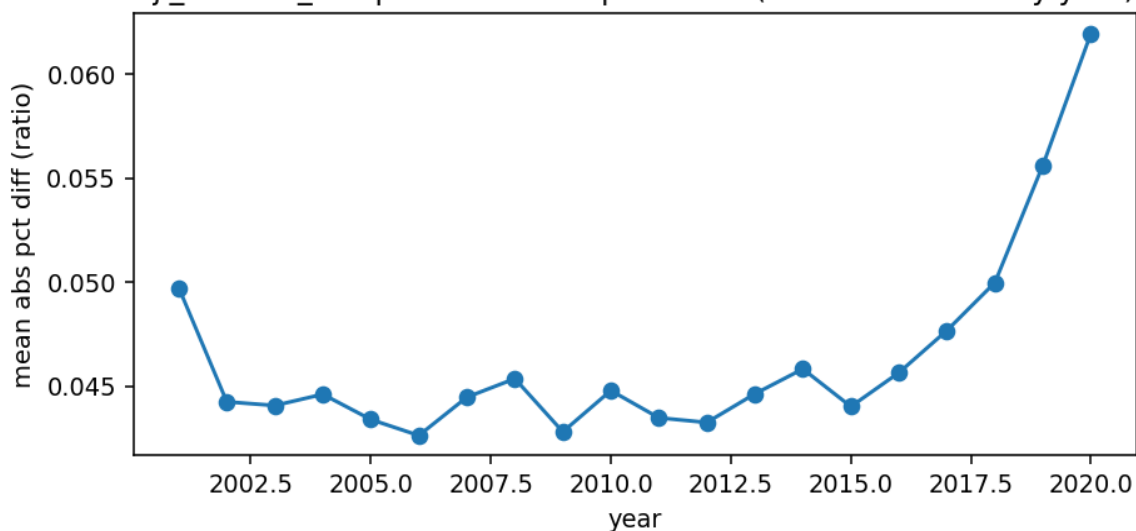
dataset_pair	metric	overlap_rows	corr	MAE	RMSE	MAPE	Median_APE	Max_APE	%>1%	%>5%	%>10%
eji_website_compensation	compensation	473349	0.8776	0.0318	0.1187	4.63%	1.59%	1466.12%	61.61%	21.26%	10.13%

Charts for primary metric: compensation

eji_website_compensation: compensation (overlap cells)



eji_website_compensation: compensation (mean abs % diff by year)



Largest revisions (top 12 by abs % diff)

region	demographic-type	demographic	year	compensation_old	compensation_new	abs_diff	abs_pct_diff
17151	Age of Business	nan	2020	0.1348320232313623	2.1116297746323345	1.9768	1466.12%
17151	Age of Business	Ages 0 to 1	2020	0.1348320232313623	2.1116297746323345	1.9768	1466.12%
17151	Age of Business	Ages 0 to 1	2019	0.1574188025041468	2.189367523201545	2.0319	1290.79%
17151	Age of Business	nan	2019	0.1574188025041468	2.189367523201545	2.0319	1290.79%
39051	Age of Business	Ages 0 to 1	2010	0.6484369679220867	8.719501396362645	8.0711	1244.70%
39051	Age of Business	nan	2010	0.6484369679220867	8.719501396362645	8.0711	1244.70%
30019	Age of Business	Ages 4 to 5	2005	0.1911902530459231	2.207376152523832	2.0162	1054.54%
8047	Age of Business	Ages 4 to 5	2020	0.5420818104440995	4.65075540832204	4.1087	757.94%
18131	Age of Business	Ages 4 to 5	2020	0.2839333099684574	2.4072679817296594	2.1233	747.83%
5129	Age of Business	Ages 4 to 5	2014	0.1566584463625154	1.0807028360049322	0.9240	589.85%
26001	Age of Business	Ages 4 to 5	2008	0.2103414359983139	1.3438993533876862	1.1336	538.91%
32011	Age of Business	Ages 2 to 3	2010	0.3835728393380099	2.4419317485184933	2.0584	536.63%

ejl_website_constancy

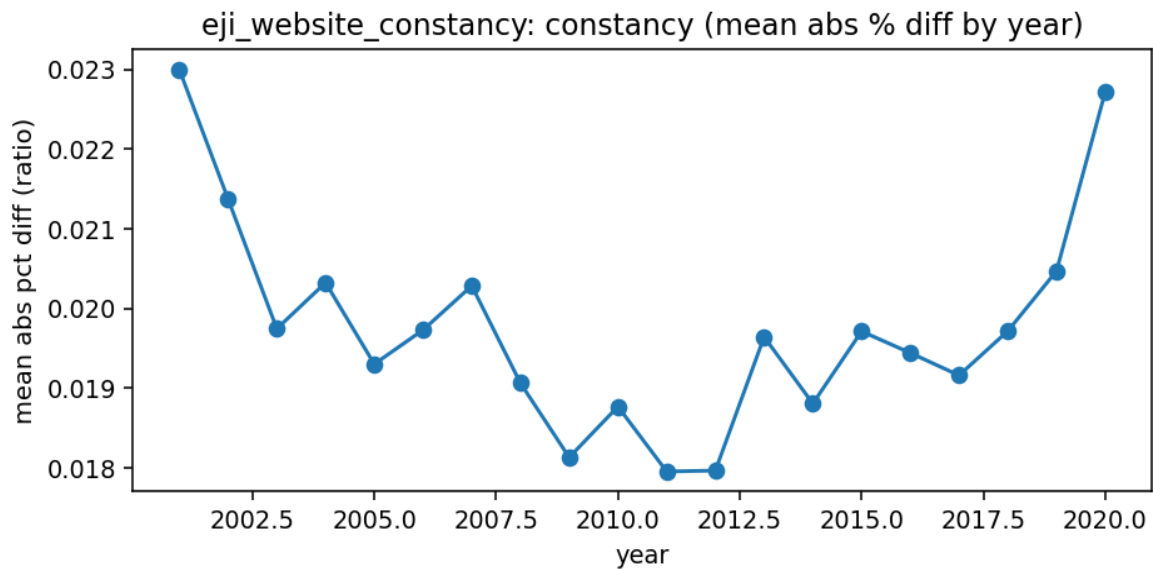
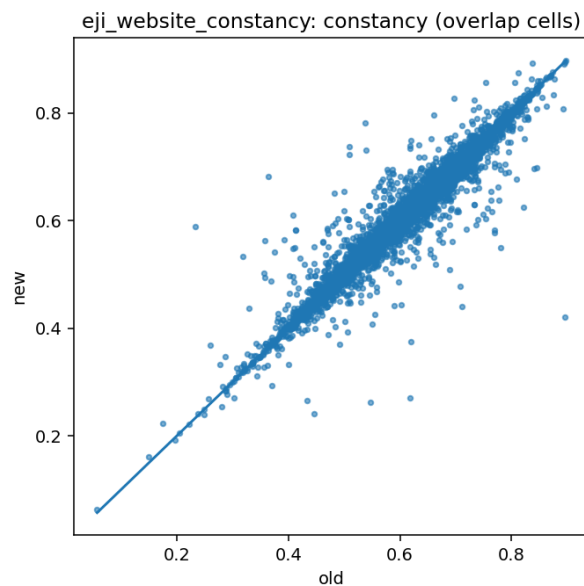
Key coverage

unique_keys_old	unique_keys_new	keys_in_both	keys_only_old	keys_only_new	old_rows	new_rows	overlap_rows
24732	24726	24360	372	366	24732	24726	487200

Similarity stats (overlap only)

dataset_pair	metric	overlap_rows	corr	MAE	RMSE	MAPE	Median_APE	Max_APE	%>1%	%>5%	%>10%
ejl_website_constancy	constancy	467401	0.9724	0.0113	0.0253	1.97%	0.75%	1429.38%	42.29%	8.79%	3.23%

Charts for primary metric: constancy



Largest revisions (top 12 by abs % diff)

region	demographic-type	demographic	year	constancy_old	constancy_new	abs_diff	abs_pct_diff
17127	Age of Business	Ages 2 to 3	2015	0.0474719101123595	0.726027397260274	0.6786	1429.38%
13269	Age of Business	Ages 0 to 1	2013	0.0634495084897229	0.5772357723577236	0.5138	809.76%
13269	Age of Business	nan	2013	0.0634495084897229	0.5772357723577236	0.5138	809.76%
29011	Age of Business	Ages 2 to 3	2014	0.0784207679826933	0.6915887850467289	0.6132	781.89%
21163	Age of Business	Ages 4 to 5	2006	0.0875912408759124	0.5979381443298969	0.5103	582.65%
13023	Age of Business	Ages 2 to 3	2014	0.1394472361809045	0.8432835820895522	0.7038	504.73%
48373	Age of Business	Ages 0 to 1	2013	0.1348829595678507	0.524788391777509	0.3899	289.07%
48373	Age of Business	nan	2013	0.1348829595678507	0.524788391777509	0.3899	289.07%
13269	Age of Business	Ages 2 to 3	2015	0.1981776765375854	0.7142857142857143	0.5161	260.43%
13269	Age of Business	Ages 6 to 10	2019	0.2061674008810572	0.7347826086956522	0.5286	256.40%
29163	Age of Business	nan	2018	0.1567080045095828	0.514760147601476	0.3581	228.48%
29163	Age of Business	Ages 0 to 1	2018	0.1567080045095828	0.514760147601476	0.3581	228.48%

ejl_website_contribution

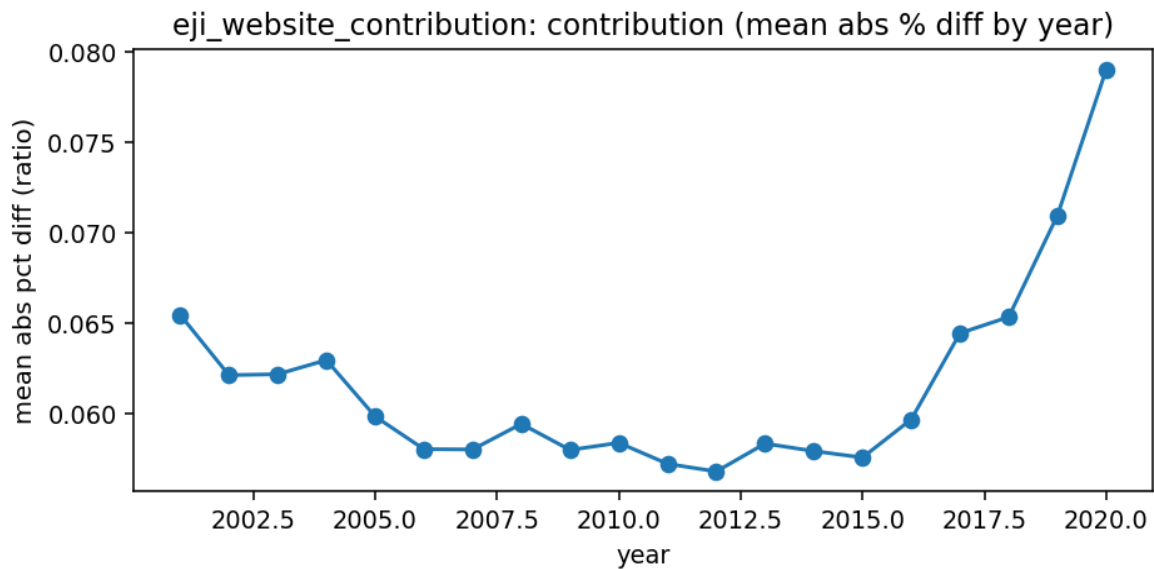
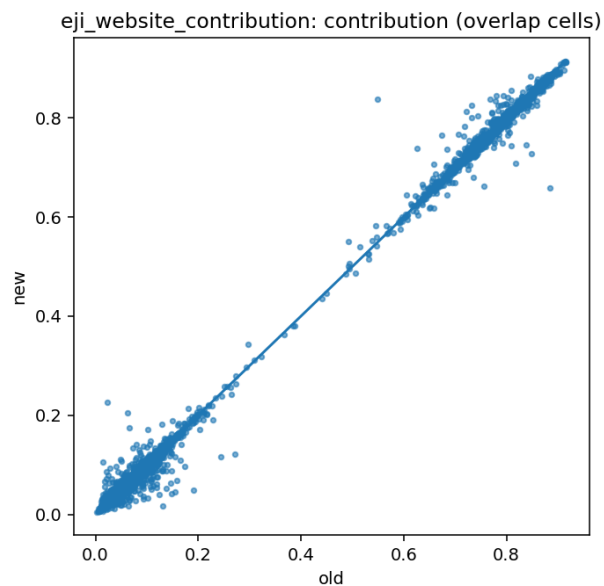
Key coverage

unique_keys_old	unique_keys_new	keys_in_both	keys_only_old	keys_only_new	old_rows	new_rows	overlap_rows
24732	24726	24360	372	366	24732	24726	487200

Similarity stats (overlap only)

dataset_pair	metric	overlap_rows	corr	MAE	RMSE	MAPE	Median_APE	Max_APE	%>1%	%>5%	%>10%
ejl_website_contribution	contribution	455176	0.9991	0.0043	0.0116	6.15%	2.59%	3193.88%	71.73%	31.62%	15.43%

Charts for primary metric: contribution



Largest revisions (top 12 by abs % diff)

region	demographic-type	demographic	year	contribution_old	contribution_new	abs_diff	abs_pct_diff
31060	Age of Business	Ages 0 to 1	2020	0.0156677079265911	0.5160759535327607	0.5004	3193.88%
31060	Age of Business	nan	2020	0.0156677079265911	0.5160759535327607	0.5004	3193.88%
35028	Age of Business	Ages 0 to 1	2020	0.0156677079265911	0.5160759535327607	0.5004	3193.88%
35028	Age of Business	nan	2020	0.0156677079265911	0.5160759535327607	0.5004	3193.88%
38081	Age of Business	Ages 0 to 1	2004	0.0173490214911754	0.4367816091954023	0.4194	2417.62%
38081	Age of Business	nan	2004	0.0173490214911754	0.4367816091954023	0.4194	2417.62%
38081	Age of Business	Ages 2 to 3	2006	0.0216273522395971	0.4480693122259065	0.4264	1971.77%
38001	Age of Business	Ages 4 to 5	2003	0.0215827338129496	0.3707144063502787	0.3491	1617.64%
46077	Age of Business	nan	2017	0.0224840078535689	0.3502659574468085	0.3278	1457.84%
46077	Age of Business	Ages 0 to 1	2017	0.0224840078535689	0.3502659574468085	0.3278	1457.84%
12039	Age of Business	Ages 2 to 3	2020	0.0195804597151411	0.3008816625636915	0.2813	1436.64%
40025	Age of Business	Ages 2 to 3	2019	0.0133475707421249	0.2003235373416015	0.1870	1400.82%

ejl_website_creation

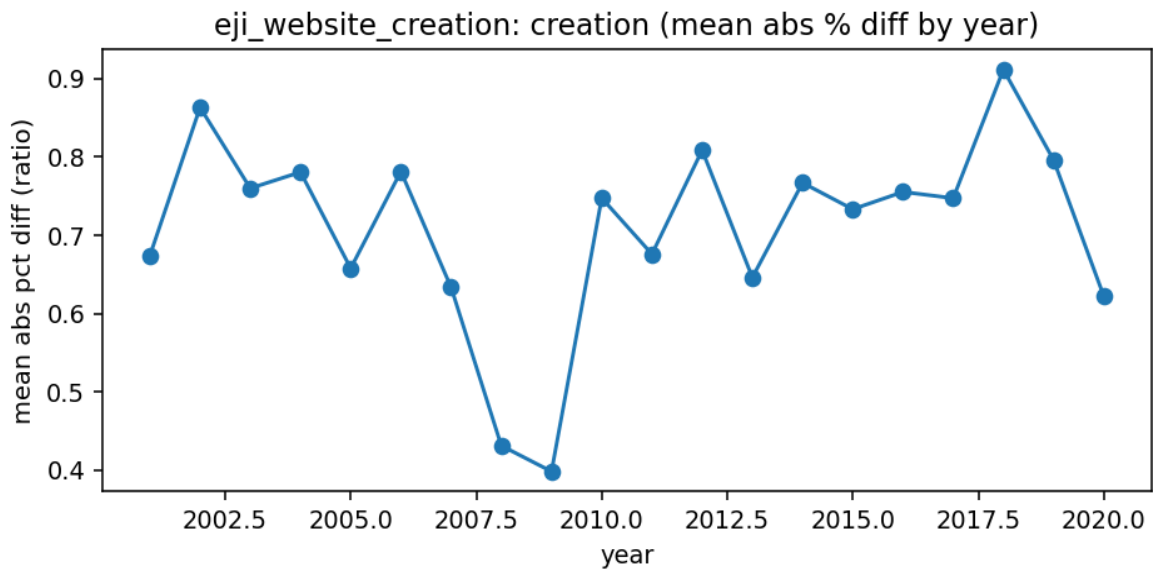
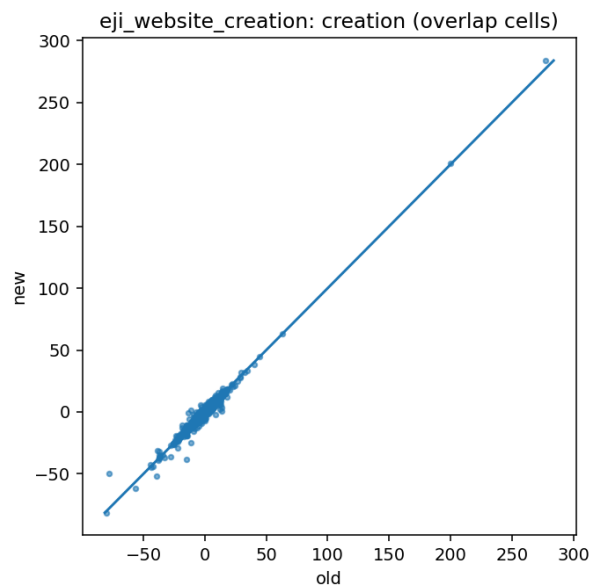
Key coverage

unique_keys_old	unique_keys_new	keys_in_both	keys_only_old	keys_only_new	old_rows	new_rows	overlap_rows
24732	24726	24360	372	366	24732	24726	487200

Similarity stats (overlap only)

dataset_pair	metric	overlap_rows	corr	MAE	RMSE	MAPE	Median_APE	Max_APE	%>1%	%>5%	%>10%
ejl_website_creation	creation	107368	0.9899	0.4424	1.2032	70.87%	12.27%	65385.37%	91.92%	69.64%	54.50%

Charts for primary metric: creation



Largest revisions (top 12 by abs % diff)

region	demographic-type	demographic	year	creation_old	creation_new	abs_diff	abs_pct_diff
53	Age of Business	Ages 2 to 3	2012	-0.0001449893563313	-0.0949468145633627	0.0948	65385.37%
22660	Age of Business	Ages 11+	2006	0.0035623572830613	-1.6778702803218946	1.6814	47200.00%
16060	Age of Business	Ages 6 to 10	2002	-0.0074883929908641	-3.307944155117135	3.3005	44074.29%
10740	Age of Business	Ages 2 to 3	2014	-0.001105536749147	0.3769451478290176	0.3781	34196.12%
47260	Age of Business	Ages 4 to 5	2001	-0.0006149161438854	0.2003212605557484	0.2009	32677.00%
12220	Age of Business	Ages 11+	2015	0.0063712935000063	1.951348580311224	1.9450	30527.20%
22180	Age of Business	Ages 11+	2014	0.0039166919619689	-1.1021577158893967	1.1061	28240.02%
18140	Age of Business	Ages 4 to 5	2018	-0.0004752413513202	0.1162440917161139	0.1167	24560.01%
29200	Age of Business	Ages 2 to 3	2001	-0.0158512937297565	-3.529554737159131	3.5137	22166.67%
46500	Age of Business	Ages 2 to 3	2013	-0.0253414763944147	5.281868969019808	5.3072	20942.78%
41940	Age of Business	Ages 11+	2004	-0.0005820667560721	0.1175774847265683	0.1182	20300.00%
17340	Age of Business	Ages 2 to 3	2003	-0.0160010240655401	-3.200204813108039	3.1842	19900.00%