

Estimating differences in duplicate end years

Thankfully, there are very few cascades in the various data design decisions made. None of the values which had different cross sectional vs. longitudinal estimates had duplicate entries. Only one value which had a duplicate start year and end year entry was also affected by having multiple observations for the same end year.

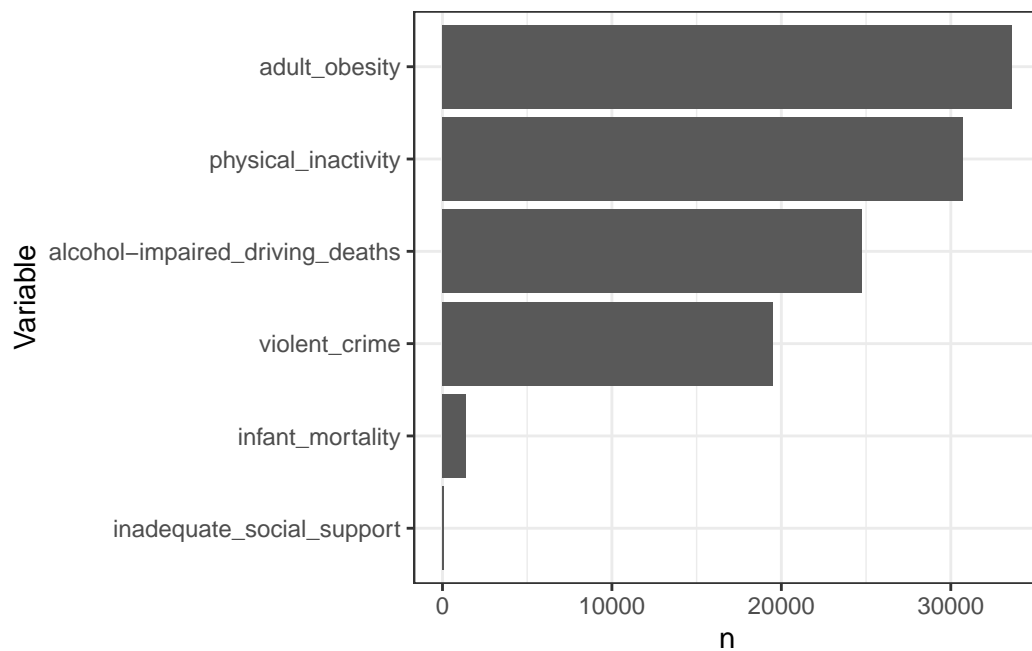
Var1	Var2	Freq
match	late	6,848
no match	late	0
match	no duplicate	3,438,354
no match	no duplicate	258,857
match	same value	461,780
no match	same value	6,185

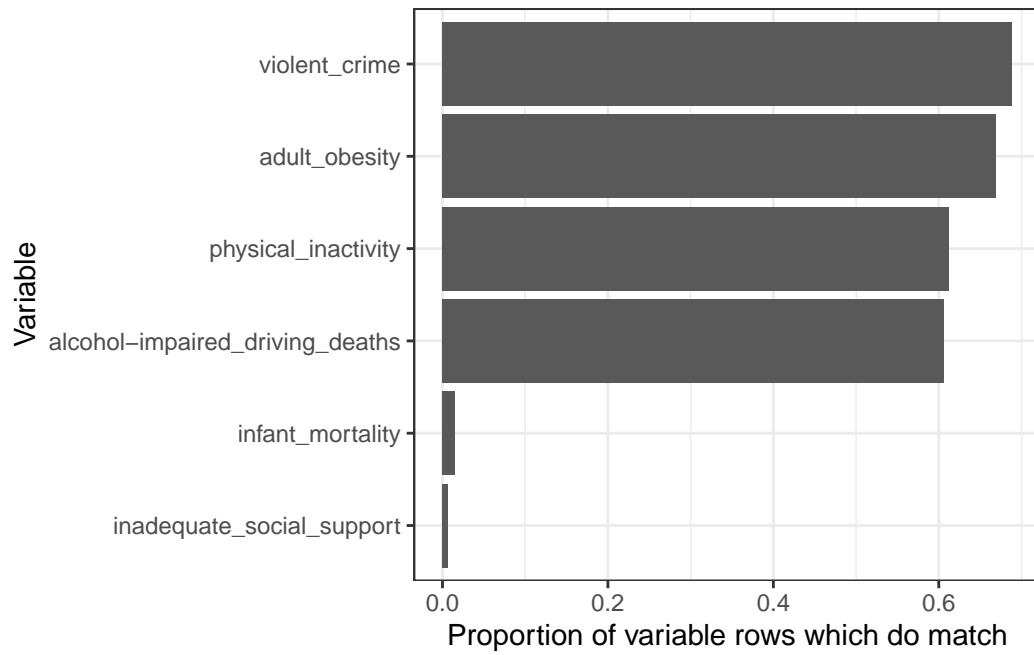
Var1	Var2	Freq
late	max end_year	1
no duplicate	max end_year	113,011
same value	max end_year	12,560
late	no duplicate	6,847

Var1	Var2	Freq
no duplicate	no duplicate	3,584,200
same value	no duplicate	455,405

Var1	Freq
min end_year	125,572
no duplicate	4,046,452

Var1	Freq
min end_year	3.009858
no duplicate	96.990142

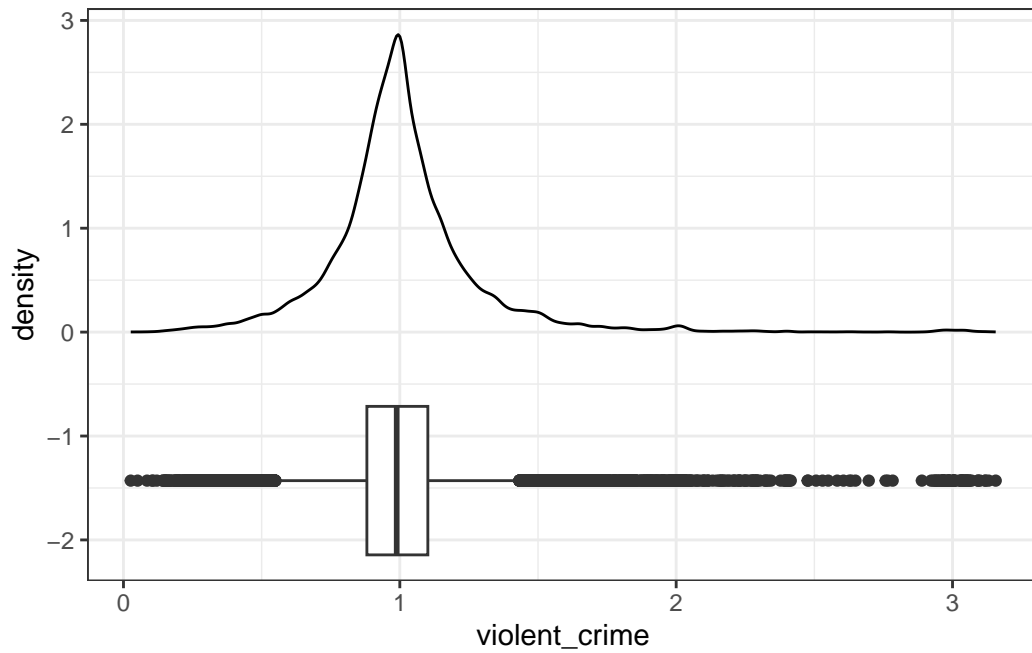




```
[[1]]
```

```
Warning: Removed 273 rows containing non-finite values (`stat_boxplot()`).
```

```
Warning: Removed 273 rows containing non-finite values (`stat_density()`).
```



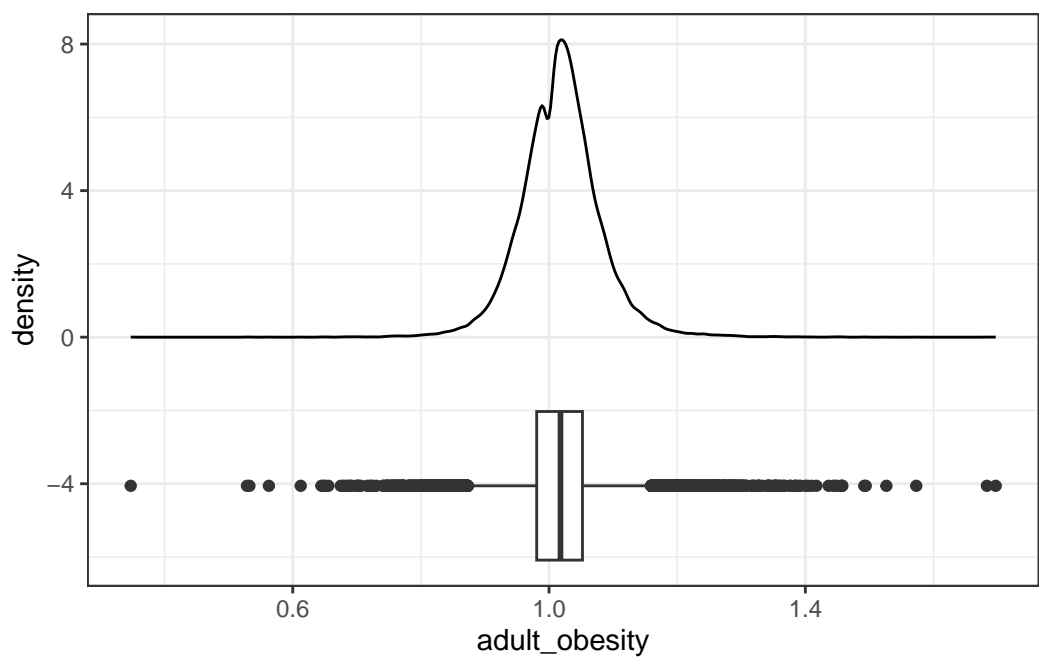
```
[[2]]
```

```
Warning: Removed 5905 rows containing non-finite values (`stat_boxplot()`).
```

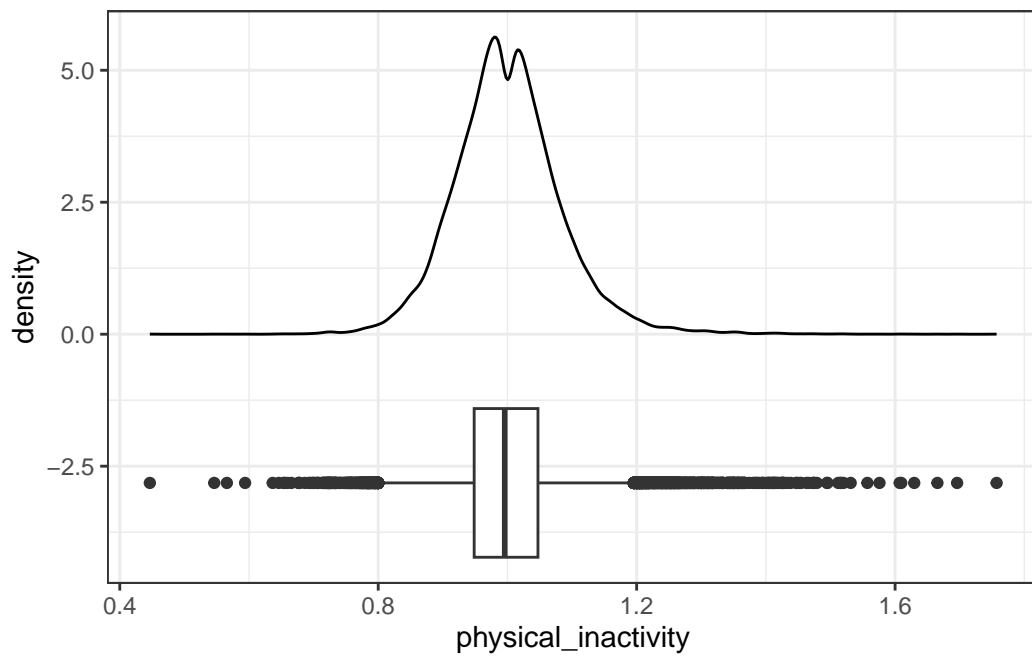
```
Warning: Removed 5905 rows containing non-finite values (`stat_density()`).
```



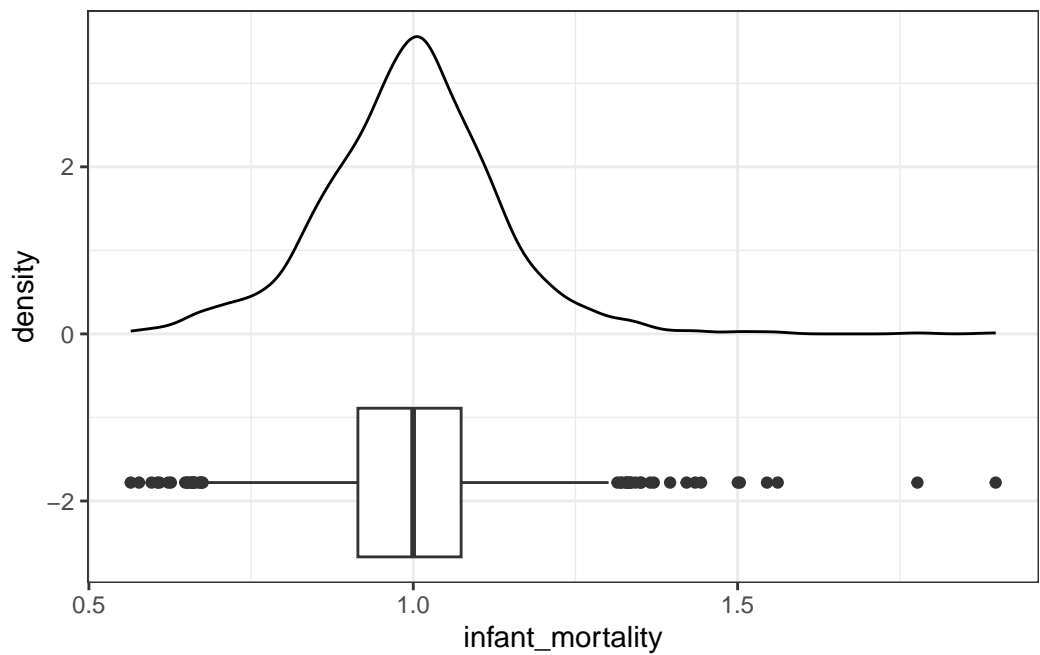
[[3]]



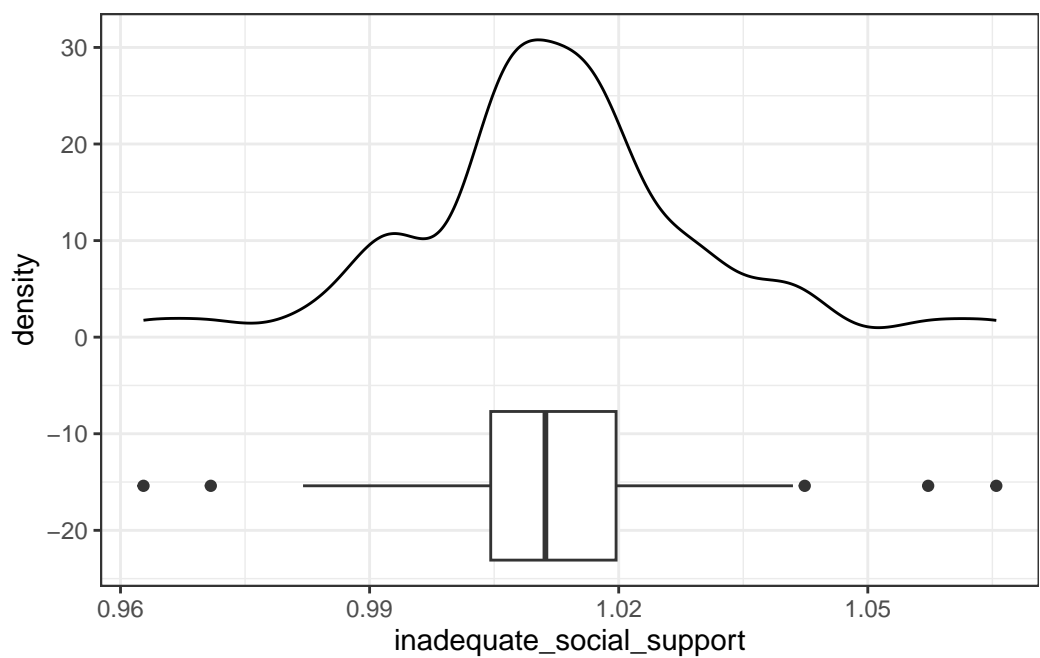
[[4]]



[[5]]



[[6]]



variable	mean	sd	min	p25	median	p75	iqr	mad	max	nr_missing	prcnt_missing	nr_rows
adult_obesity	1.018	0.064	0.347	0.981	1.018	1.052	0.071	0.053	1.697	0	0.000	33,639
alcohol- impaired_driving_deaths	1.280	0.892	0.136	0.794	1.083	1.500	0.706	0.494	29.000	5,905	0.238	24,763
inadequate_social_support	1.012	0.018	0.963	1.005	1.011	1.020	0.015	0.010	1.065	0	0.000	61
infant_mortality	0.996	0.136	0.565	0.915	1.000	1.074	0.159	0.118	1.898	0	0.000	1,371
physical_inactivity	1.001	0.084	0.447	0.948	0.996	1.047	0.099	0.073	1.757	0	0.000	30,741
violent_crime	1.008	0.278	0.026	0.881	0.989	1.101	0.221	0.163	3.156	273	0.014	19,510