Product Summary

GB Startup Index is a data product produced by Datadaptive from statistical analysis of records in the Companies House register of companies. The Index is designed as a basis for geographic comparison of levels of company formations over the most recent twelve month period. The Index is recalculated once a month. June 2017 is the ninth release.

The Index uses three measures:

- the number of new company registrations in the past twelve months per head of population,
- the number of new company registrations in the past twelve months as a percentage of all live company registrations in the area, and
- the median average date of incorporation of all live company registrations in the area.

Each measure is normalised via a modified Z-score onto a point scale from 0 to 100. The three measures are then combined to produce a weighted composite index of startup activity for each local authority district and Westminster parliamentary constituency in England, Scotland and Wales.

The data product includes the above measures at various stages of calculation along with the composite measure, as a table for each of the two geographies. Selected variables are also provided as attributes attached to spatial boundary datasets for the two geographies, suitable for thematic mapping.

Data quality statement

The primary basis of calculation of the measures in the Index is records of live company registrations as published by Companies House in a data release of 01/06/2017 with a cut-off date at the end of May 2017. For purposes of these calculations Datadaptive discarded all records with a country of origin outside the

UK and all records with a postcode location outside Great Britain. Records were then matched by postcode to local authority districts and Westminster parliamentary constituencies using ONS postcode data. Slightly less than 1% of Companies House records for companies based in Great Britain were discarded because the postcode data in those records did not automatically match any live or expired postcode in the ONS dataset.

Users should also bear in mind that geolocation of the Companies House records, and therefore the geographic comparisons enabled by the measures in the Index, is based on registration addresses. Registration addresses do not necessarily correspond to physical locations at which companies conduct business. In some cases the registration address will be that of an accountant or other proxy.

Although the design of the Index measures does include fencing of outliers, results for some geographic areas may be skewed by the activities of formation agents and similar businesses involved in the bulk registration of companies. Please see Appendix A for further details on calculation of the Index measures.

Datadaptive has applied professional judgement to the selection and design of the Index measures and it is a matter for the user to decide whether those measures are suitable for the user's purposes.

The data in this product has been compiled and/or calculated from data published by ONS for re-use under an open licence. Datadaptive has taken reasonable care in production of the data. However the accuracy and currency of the data and calculations is partly dependent on the input sources.

Data formats

Tabular data is normally supplied in CSV format. Spatial data is normally supplied in unstyled ESRI Shapefile format with the National Grid (OSGB36) projection. However we are happy to prepare the data in other suitable formats.

Licensing

Supply of the GB Startup Index data product is subject to our standard notice for supply of added value open licensed data products. Once supplied the data may be re-used under the Creative Commons Attribution License 4.0 (CC-BY).

CC-BY is an open data licence that allows almost unrestricted re-use of the data. CC-BY is perpetual, royalty-free, non-exclusive, and interoperable with the [Open Government Licence](#).

Attribution

Following is an attribution statement that acknowledges Datadaptive's intellectual property rights as well as those of the third-party providers of open data used in the data product. You must include this statement if you publish or redistribute the GB Startup Index data.

Pricing

£850 one-off supply fee

Contact details

Email: info@datadaptive.com

Website: www.datadaptive.com/

File Specifications

GB Startup Index for Local Authority Districts: June 2017

File name: STARTUP_INDEX_GB_LAUA_201706

Record count: 380

Geographic coverage: England, Scotland and Wales

File format: CSV

This file is a table of all local authority district and unitary councils in Britain, each attributed with measures for comparing levels of new business registrations including a range of variables and intermediate calculations used to produce the composite index.

Following is a table of the fields provided for each district or unitary.

Field	Type	Description
LAUACD	Text	ONS code
LAUANM	Text	ONS name
LAUANMW	Text	ONS Welsh name
COMPCOUNTT	Integer	Total number of live companies registered at postcodes in area
COMPCOUNT	Integer	Number of live companies registered at postcodes in area within the past 12 months (June 2016 - May 2017)
CMEDAGE	Date	Median average date of incorporation of live companies registered at postcodes in area
COMPNRAT	Decimal	New companies as percentage of total companies (COMPCOUNT / COMPCOUNTT * 100)
POPCOUNT	Integer	Resident population in area (Mid-2015 estimate)
POPCOMP	Decimal	Resident population per number of new companies (POPCOUNT / COMPCOUNT)
COMPPOP	Decimal	New companies per head of population (COMPCOUNT/POPCOUNT)
COMPRANK	Integer	Rank of COMPCOUNT (from 1 to 380 with 1 corresponding to the area with the highest value of COMPCOUNT)
CPRANK	Integer	Rank of COMPPOP (from 1 to 380 with 1 corresponding to the area with the highest value of COMPPOP)
CPZSCORE	Decimal	Z-score of raw score COMPPOP
CPMZSCORE	Decimal	Modified Z-score of raw score COMPPOP
CPMZSCALE	Decimal	Modified Z-score CPMZSCORE transformed to a scale from 0 to 100 with outliers fenced
CNRRANK	Integer	Rank of COMPNRAT (from 1 to 380 with 1 corresponding to the area with the highest value of COMPNRAT)
CNRZSCORE	Decimal	Z-score of raw score COMPNRAT
CNRMZSCORE	Decimal	Modified Z-score of raw score COMPNRAT

CNRMZSCALE	Decimal	Modified Z-score CNRMZSCORE transformed to a scale from 0 to 100 with outliers fenced
CMARANK	Integer	Rank of CMEDAGE (from 1 to 380 with 1 corresponding to the area with the most recent CMEDAGE date)
CMAZSCORE	Decimal	Z-score of date (as raw score) CMEDAGE
CMAMZSCORE	Decimal	Modified Z-score of date (as raw score) CMEDAGE
CMAMZSCALE	Decimal	Modified Z-score CMAMZSCORE transformed to a scale from 0 to 100 with outliers fenced
IND12M	Decimal	Composite index on a scale from 0 to 100 calculated by weighted averaging of CPMZSCALE, CNRMZSCALE and CMAMZSCALE
IND12MRANK	Integer	Rank of IND12M (from 1 to 380 with 1 corresponding to the area with the highest IND12M score)

Please see Appendix A for additional technical notes on the calculation of some of the above fields.

GB Startup Index for Westminster Parliamentary Constituencies: June 2017

File name: STARTUP_INDEX_GB_PCON_201706
Record count: 632
Geographic coverage: England, Scotland and Wales
File format: CSV

This file is a table of all Westminster parliamentary constituencies in Britain, each attributed with measures for comparing levels of new business registrations including a range of variables and intermediate calculations used to produce the composite index.

Following is a table of the fields provided for each constituency.

Field	Type	Description
PCONCD	Text	ONS code
PCONNM	Text	ONS name
PCONNMW	Text	ONS Welsh name
COMPCOUNTT	Integer	Total number of live companies registered at postcodes in area
COMPCOUNT	Integer	Number of live companies registered at postcodes in area within the past 12 months (June 2016 - May 2017)
CMEDAGE	Date	Median average date of incorporation of live companies registered at postcodes in area
COMPNRAT	Decimal	New companies as percentage of total companies (COMPCOUNT / COMPCOUNTT * 100)
POPCOUNT	Integer	Resident population in area (Mid-2015 estimate)
POPCOMP	Decimal	Resident population per number of new companies (POPCOUNT / COMPCOUNT)
COMPPOP	Decimal	New companies per head of population (COMPCOUNT/POPCOUNT)
COMPRANK	Integer	Rank of COMPCOUNT (from 1 to 632 with 1 corresponding to the area with the highest value of COMPCOUNT)
CPRANK	Integer	Rank of COMPPOP (from 1 to 632 with 1 corresponding to the area with the highest value of COMPPOP)
CPZSCORE	Decimal	Z-score of raw score COMPPOP
CPMZSCORE	Decimal	Modified Z-score of raw score COMPPOP
CPMZSCALE	Decimal	Modified Z-score CPMZSCORE transformed to a scale from 0 to 100 with outliers fenced
CNRRANK	Integer	Rank of COMPNRAT (from 1 to 632 with 1 corresponding to the area with the highest value of COMPNRAT)
CNRZSCORE	Decimal	Z-score of raw score COMPNRAT
CNRMZSCORE	Decimal	Modified Z-score of raw score COMPNRAT

CNRMZSCALE	Decimal	Modified Z-score CNRMZSCORE transformed to a scale from 0 to 100 with outliers fenced
CMARANK	Integer	Rank of CMEDAGE (from 1 to 632 with 1 corresponding to the area with the most recent CMEDAGE date)
CMAZSCORE	Decimal	Z-score of date (as raw score) CMEDAGE
CMAMZSCORE	Decimal	Modified Z-score of date (as raw score) CMEDAGE
CMAMZSCALE	Decimal	Modified Z-score CMAMZSCORE transformed to a scale from 0 to 100 with outliers fenced
IND12M	Decimal	Composite index on a scale from 0 to 100 calculated by weighted averaging of CPMZSCALE, CNRMZSCALE and CMAMZSCALE
IND12MRANK	Integer	Rank of IND12M (from 1 to 632 with 1 corresponding to the area with the highest IND12M score)

Please see Appendix A for additional technical notes on the calculation of some of the above fields.

Local Authority Districts Boundaries with GB Startup Index Attributes: June 2017

File name: STARTUP_INDEX_GB_LAUA_201706 and
                               STARTUP_INDEX_GB_LAUA_201706_UG
Record count: 380
Geographic coverage: England, Scotland and Wales
File format: SHP

This is a set of two spatial datasets each of which contains polygons for all local authority district and unitary councils in Britain attributed with a selection of measures for comparing levels of new business registrations, including the composite index and population denominator.

File STARTUP_INDEX_GB_LAUA_201706 contains full extent boundaries and file STARTUP_INDEX_GB_LAUA_201706_UG contains boundaries that have been clipped to the coastline and generalised to 500 metres.

Following is a table of the attributes provided for each district or unitary.

Field	Type	Description
LAUACD	Text	ONS code
LAUANM	Text	ONS name
LAUANMW	Text	ONS Welsh name
POPCOUNT	Integer	Resident population in area (Mid-2015 estimate)
COMPCOUNTT	Integer	Total number of live companies registered at postcodes in area
COMPCOUNT	Integer	Number of live companies registered at postcodes in area within the past 12 months (June 2016 - May 2017)
CMEDAGE	Date	Median average date of incorporation of live companies registered at postcodes in area
COMPNRAT	Decimal	New companies as percentage of total companies (COMPCOUNT / COMPCOUNTT * 100)
COMPPOP	Decimal	New companies per head of population (COMPCOUNT/POPCOUNT)
IND12M	Decimal	Composite index on a scale from 0 to 100 calculated by weighted averaging of scales derived via modified Z-scores from COMPPOP, COMPNRAT and CMEDAGE
IND12MRANK	Integer	Rank of IND12M (from 1 to 380 with 1 corresponding to the area with the highest IND12M score)

Please see Appendix A for additional technical notes on the calculation of some of the above fields.

Westminster Parliamentary Constituency Boundaries with GB Startup Index Attributes: June 2017

File name: STARTUP_INDEX_GB_PCON_201706 and STARTUP_INDEX_GB_PCON_201706_UG

Record count: 632

Geographic coverage: England, Scotland and Wales

File format: SHP

This is a set of two spatial datasets each of which contains polygons for all Westminster parliamentary constituencies in Britain attributed with a selection of measures for comparing levels of new business registrations, including the composite index and population denominator.

File STARTUP_INDEX_GB_PCON_201706 contains full extent boundaries and file STARTUP_INDEX_GB_PCON_201706_UG contains boundaries that have been clipped to the coastline and generalised to 500 metres.

Following is a table of the attributes provided for each constituency.

Field	Type	Description
PCONCD	Text	ONS code
PCONNM	Text	ONS name
PCONNMW	Text	ONS Welsh name
POPCOUNT	Integer	Resident population in area (Mid-2015 estimate)
COMPCOUNTT	Integer	Total number of live companies registered at postcodes in area
COMPCOUNT	Integer	Number of live companies registered at postcodes in area within the past 12 months (June 2016 - May 2017)
CMEDAGE	Date	Median average date of incorporation of live companies registered at postcodes in area
COMPNRAT	Decimal	New companies as percentage of total companies (COMPCOUNT / COMPCOUNTT * 100)
COMPPOP	Decimal	New companies per head of population (COMPCOUNT/POPCOUNT)
IND12M	Decimal	Composite index on a scale from 0 to 100 calculated by weighted averaging of scales derived via modified Z-scores from COMPPOP, COMPNRAT and CMEDAGE
IND12MRANK	Integer	Rank of IND12M (from 1 to 632 with 1 corresponding to the area with the highest IND12M score)

Please see Appendix A for additional technical notes on the calculation of some of the above fields.

Appendix A: Calculation of Index Measures

Following are additional technical notes on the calculations used to produce some of the values in the Index datasets. (In points 1, 2 and 4 below, "population" is used with its generic statistical meaning rather than in reference to the resident population of an area.)

1. Z-scores (fields CPZSCORE, CNRZSCORE and CMAZSCORE) were calculated by subtracting the mean of the population from the raw score and then dividing by the standard deviation of the population.

2. Modified Z-scores (fields CPMZSCORE, CNRMZSCORE and CMAMZSCORE) were calculated by subtracting the median of the population from the raw score, multiplying by a constant of 0.6745, and then dividing by the median absolute deviation of the population.

3. Scales for individual measures (fields CPMZSCALE, CNRMZSCALE and CMAMZSCALE) were calculated as linear transformations of the modified Z-scores to a range from 0 to 100 after fencing of any modified Z-score with an absolute value of 3.5 or higher. In other words any modified z-score with a value of 3.5 or higher or a value of -3.5 or lower was treated as an outlier and assigned a scale point of 100 or 0 respectively.

4. Modified Z-scores were used in preference to Z-scores in the design of the Index measures in order to dampen the unrepresentative influence on the results of a small number of formation agents operating in particular geographic locations. This, along with the fencing of outliers, has the effect of elongating the spread of points along the 0 to 100 scales in order to make it easier to differentiate between members (i.e. geographic locations) in the mid-range of the population.

5. The composite index (field IND12M) is a weighted average of the three individual 0 to 100 scales (fields CPMZSCALE, CNRMZSCALE and CMAMZSCALE). The scale for the number of new company registrations in the past twelve months per head of resident population (field CPMZSCALE) has been assigned a weighting of 50%. The scales for the number of new company registrations in the past twelve months as a percentage of all live company registrations in the area (field CNRMZSCALE) and the median average date of incorporation of all live company registrations in the area (field CMAMZSCALE) have each been assigned a weighting of 25%.