Package ‘SGPdata’

October 12, 2022

Type Package

Title Exemplar Data Sets for Student Growth Percentiles (SGP) Analyses

Version 26.0-0.0

Date 2022-5-26

Maintainer Damian W. Betebenner <dbetebenner@nciea.org>

Depends R (>= 4.0)

Suggests SGP (>= 1.9-0.0), knitr, rmarkdown

Imports crayon, data.table (>= 1.14.0)

Description Data sets utilized by the ‘SGP’ package as exemplars for users to conduct their own student growth percentiles (SGP) analyses.

URL https://CenterForAssessment.github.io/SGPdata/,
https://github.com/CenterForAssessment/SGPdata/,
https://cran.r-project.org/package=SGPdata

BugReports https://github.com/CenterForAssessment/SGPdata/issues/

VignetteBuilder knitr

LazyLoad Yes

LazyData Yes

LazyDataCompression xz

License GPL-3

NeedsCompilation no

Author Damian W. Betebenner [aut, cre],
Adam R. Van Iwaarden [aut],
Ben Domingue [aut]

Repository CRAN

Date/Publication 2022-05-27 09:30:10 UTC
### R topics documented:

<table>
<thead>
<tr>
<th>Package</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGPdata-package</td>
<td>2</td>
</tr>
<tr>
<td>sgpData</td>
<td>2</td>
</tr>
<tr>
<td>sgpData_INSTRUCTOR_NUMBER</td>
<td>3</td>
</tr>
<tr>
<td>sgpData_LONG</td>
<td>4</td>
</tr>
<tr>
<td>sgpData_LONG_COVID</td>
<td>6</td>
</tr>
<tr>
<td>sgpData_LONG_COVID</td>
<td>7</td>
</tr>
</tbody>
</table>

### Index

9

---

### SGPdata-package

**Exemplar Data Sets for Student Growth Percentile (SGP) Analyses**

### Description

The 'SGPdata' package contains five exemplar data sets utilized by the 'SGP' package for testing and examples.

### Details

<table>
<thead>
<tr>
<th>Package:</th>
<th>SGPdata</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type:</td>
<td>Package</td>
</tr>
<tr>
<td>Version:</td>
<td>26.0-0.0</td>
</tr>
<tr>
<td>Date:</td>
<td>2022-5-26</td>
</tr>
<tr>
<td>License:</td>
<td>GPL-3</td>
</tr>
<tr>
<td>LazyLoad:</td>
<td>Yes</td>
</tr>
<tr>
<td>LazyData:</td>
<td>Yes</td>
</tr>
</tbody>
</table>

---

### sgpData

**Longitudinal Student Assessment Data**

### Description

A dataset comprising a subset of five years of state assessment reading data suitable for student growth percentile and percentile growth projection/trajectory analyses. The dataset is used in examples provided in the documentation for the `studentGrowthPercentiles` and `studentGrowthProjections` functions, and serves as an exemplar for user construction of their own datasets for growth percentile analyses.
Format

A data frame of student level observations for five years on the following variables.

- **ID**  Student ID
- **GRADE_2018**  Student Grade Tested 2018, possibly missing
- **GRADE_2019**  Student Grade Tested 2019, possibly missing
- **GRADE_2020**  Student Grade Tested 2020, possibly missing
- **GRADE_2021**  Student Grade Tested 2021, possibly missing
- **GRADE_2022**  Student Grade Tested 2022, possibly missing
- **SS_2018**  Student Scale Score 2018, possibly missing
- **SS_2019**  Student Scale Score 2019, possibly missing
- **SS_2020**  Student Scale Score 2020, possibly missing
- **SS_2021**  Student Scale Score 2021, possibly missing
- **SS_2022**  Student Scale Score 2022, possibly missing

Note

All datasets used with `studentGrowthPercentiles` and `studentGrowthProjections` must be specifically formatted as wide format files. The first variable/column is the student ID variable. The next set of columns provide the grade of the student across all the years provided in the dataset (possibly missing). The last set of columns provide the scales scores of the student in the respective grades. Multi-year operational analyses benefit from putting data in long format. See `sgpData_LONG` and associated documentation for a comprehensive account.

Source

Anonymized student level state assessment data

See Also

- `sgpData_LONG`
sgpData_LONG

Format
A data frame of student level observations for five years and two content areas for the following variables.

ID Unique Student Identification Number
CONTENT_AREA Content area for student/teacher record (Reading or Mathematics)
YEAR Year for student/teacher record
INSTRUCTOR_NUMBER Unique instructor number identifier
INSTRUCTOR_LAST_NAME Instructor last name
INSTRUCTOR_FIRST_NAME Instructor first name
INSTRUCTOR_WEIGHT Proportional weight associated with student/teacher record
INSTRUCTOR_ENROLLMENT_STATUS Indicator of full enrollment status of student with teacher

Source
Anonymized student/teacher lookup table in LONG format

See Also
sgpData_LONG

---

sgpData_LONG Longitudinal Student Assessment Data in LONG format

Description
A dataset comprising a subset of five years of state assessment reading and mathematics data suitable for student growth percentile (SGP) and percentile growth projection/trajectory analyses. The dataset is in LONG format with each record representing a unique student, by year by content area combination. The example data is used in examples and illustration in the enclosed documentation to show how SGP can be operationalized across years with the results being used to produce a variety of high quality visualizations both at the aggregate (e.g., school) and individual level.

Format
A data frame of student level observations for five years and two content areas for the following variables.

ID Unique Student Identification Number
LAST_NAME Student last name
FIRST_NAME Student first name
CONTENT_AREA Content area for student observation (Reading or Mathematics)
YEAR Year for student observation
GRADE Grade level of student observation
**SGPData_LONG**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCALE_SCORE</td>
<td>Student Scale Score</td>
</tr>
<tr>
<td>ACHIEVEMENT_LEVEL</td>
<td>Achievement level associated with student scale score</td>
</tr>
<tr>
<td>GENDER</td>
<td>Student gender</td>
</tr>
<tr>
<td>ETHNICITY</td>
<td>Student ethnicity</td>
</tr>
<tr>
<td>FREE_REDUCED_LUNCH_STATUS</td>
<td>Student free/reduced lunch status</td>
</tr>
<tr>
<td>ELL_STATUS</td>
<td>Student English Language Learner status</td>
</tr>
<tr>
<td>IEP_STATUS</td>
<td>Student Individual Education Plan status</td>
</tr>
<tr>
<td>GIFTED_AND_TALENTED_PROGRAM_STATUS</td>
<td>Student Gifted and Talented Program status</td>
</tr>
<tr>
<td>SCHOOL_NUMBER</td>
<td>School number associated with student record</td>
</tr>
<tr>
<td>SCHOOL_NAME</td>
<td>School name associated with school number and student record</td>
</tr>
<tr>
<td>EMH_LEVEL</td>
<td>Elementary, Middle, High school indicator for school attended</td>
</tr>
<tr>
<td>DISTRICT_NUMBER</td>
<td>District number associated with student record</td>
</tr>
<tr>
<td>DISTRICT_NAME</td>
<td>District name associated with school number and student record</td>
</tr>
<tr>
<td>SCHOOL_ENROLLMENT_STATUS</td>
<td>Indicator of full academic year student enrollment in school</td>
</tr>
<tr>
<td>DISTRICT_ENROLLMENT_STATUS</td>
<td>Indicator of full academic year student enrollment in school</td>
</tr>
<tr>
<td>STATE_ENROLLMENT_STATUS</td>
<td>Indicator of full academic year student enrollment in school</td>
</tr>
<tr>
<td>VALID_CASE</td>
<td>Indicator of whether the case is valid or invalid</td>
</tr>
</tbody>
</table>

**Note**

All datasets used with studentGrowthPercentiles and studentGrowthProjections must be specifically formatted as wide format files. The first variable/column is the student ID variable. The next set of columns provide the grade of the student across all the years provided in the dataset (possibly missing). The last set of columns provide the scales scores of the student in the respective grades. Operational users of this package are advised to use data formatted similar to the sgpData_LONG with the higher level functions prepareSGP, analyzeSGP, combineSGP, summarizeSGP, visualizeSGP, and outputSGP.

**Source**

Anonymized student level state assessment data in LONG format

**See Also**

sgpData
sgpData_LONG_COVID

Longitudinal Student Assessment Data in LONG format with COVID-19 2020 interruption

Description

A dataset comprising a subset of seven years of state assessment ELA and mathematics data across an 8 year span suitable for student growth percentile (SGP) and percentile growth projection/trajectory analyses. The dataset is missing 2020 data to help users model COVID related interruptions to student growth. The dataset is in LONG format with each record representing a unique student, by year by content area combination. The example data is used in examples and illustration in the enclosed documentation to show how SGPs can be operationalized across years with the results being used to produce a variety of high quality visualizations both at the aggregate (e.g., school) and individual level.

Format

A data frame of student level observations for seven years (across an 8 year span) and two content areas for the following variables.

- **VALID_CASE**  Indicator of whether the case is valid or invalid
- **CONTENT_AREA**  Content area for student observation (ELA or Mathematics)
- **YEAR**  Year for student observation
- **ID**  Unique Student Identification Number
- **IEP_STATUS**  Student Individual Education Plan status
- **GRADE**  Grade level of student observation
- **SCALE_SCORE**  Student Scale Score
- **SCALE_SCORE_without_COVID_IMPACT**  Student Scale Score without COVID impact. Original scale score
- **ACHIEVEMENT_LEVEL**  Achievement level associated with student scale score
- **ETHNICITY**  Student ethnicity
- **FREE_REduced_LUNCH_STATUS**  Student free/reduced lunch status
- **ELL_STATUS**  Student English Language Learner status
- **GENDER**  Student gender
- **SCHOOL_NUMBER**  School number associated with student record
- **SCHOOL_NAME**  School name associated with school number and student record
- **DISTRICT_NUMBER**  District number associated with student record
- **DISTRICT_NAME**  District name associated with school number and student record
Note

All datasets used with studentGrowthPercentiles and studentGrowthProjections must be specifically formatted as *wide* format files. The first variable/column is the student ID variable. The next set of columns provide the grade of the student across all the years provided in the dataset (possibly missing). The last set of columns provide the scales scores of the student in the respective grades. Operational users of this package are advised to use data formatted similar to the sgpData_LONG_COVID with the higher level functions prepareSGP, analyzeSGP, combineSGP, summarizeSGP, visualizeSGP, and outputSGP.

Source

Anonymized and altered student level state assessment data in LONG format

See Also

- sgpData_LONG

---

## Description

A dataset comprising a subset of 3 years (and time periods) of interim assessment early literacy, reading and mathematics data suitable for time dependent student growth percentile (SGPt) and percentile growth projection/trajectory analyses. The dataset is in LONG format with each record representing a unique student, by year/time period by content area combination. The example data is used in examples and illustrations in the enclosed documentation to show how SGPts can be operationalized across years with the results being used to produce a variety of high quality visualizations both at the aggregate (e.g., school) and individual level.

## Format

A data frame of student level observations for five years and two content areas for the following variables.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>COUNTRY</td>
<td>Variable indicating the COUNTRY associated with test record.</td>
</tr>
<tr>
<td>STATE</td>
<td>State associated with the test record.</td>
</tr>
<tr>
<td>VALID_CASE</td>
<td>VALID_CASE variable.</td>
</tr>
<tr>
<td>CONTENT_AREA</td>
<td>Variable indicating content area associated with test record.</td>
</tr>
<tr>
<td>YEAR</td>
<td>Variable indicating the year associated with the test record.</td>
</tr>
<tr>
<td>DATE</td>
<td>Variable indicating the year associated with the test record: Must be in YYYY-MM-DD format.</td>
</tr>
<tr>
<td>ID</td>
<td>Unique student identifier associated with test record.</td>
</tr>
<tr>
<td>GRADE</td>
<td>Variable indicating the grade associated with the test record.</td>
</tr>
</tbody>
</table>
SCALE_SCORE  Variable providing the scale score associated with the test record.
SCALE_SCORE_RASCH  Variable providing the Rasch score associated with the test record.
SEM  Variable indicating the standard error of measurement associated with the scale score.
ACHIEVEMENT_LEVEL  Variable indicating the achievement level associated with the scale score.

Source
Anonymized student level interim assessment data in LONG format

See Also
sgpData_LONG
Index

* datasets
  sgpData, 2
  sgpData_INSTRUCTOR_NUMBER, 3
  sgpData_LONG, 4
  sgpData_LONG_COVID, 6
  sgptData_LONG, 7

* package
  SGPdata-package, 2
  SGPdata (SGPdata-package), 2
  sgpData, 2, 5
  SGPdata-package, 2
  sgpData_INSTRUCTOR_NUMBER, 3
  sgpData_LONG, 3, 4, 7, 8
  sgpData_LONG_COVID, 6
  sgptData_LONG, 7