CS 2334: Project 3
Java Collections Framework

Project 2 Lessons

Project 2 Lessons

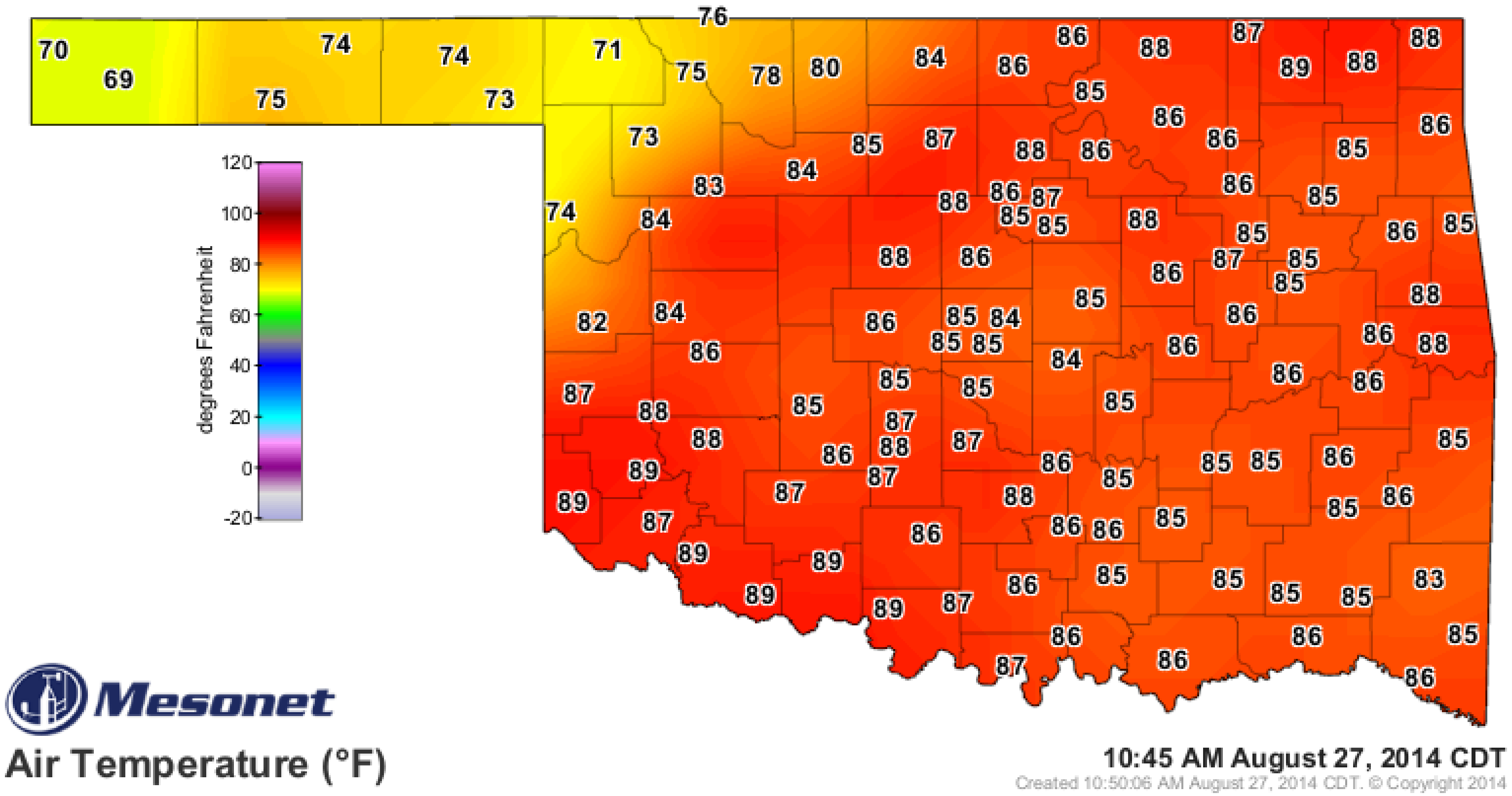
- Class hierarchies
- Code reuse through abstract classes
- Computing statistics in the presence of invalid data

- Samples: should only ask what the value is if we know that it is valid

Project 3

Expanded Mesonet data:

- Many variables
- Many stations
- Your program won’t know which variables and stations there are until it is executed!



A Few of our Stations…. (138 in all)

```
Station ID   Name                       City
----------   ----                       ----
ACME         Acme                       Rush Springs
ADAX         Ada                        Ada
ALTU         Altus                      Altus
ALV2         Alva                       Alva
ALVA         Alva                       Alva
ANT2         Antlers                    Antlers
ANTL         Antlers                    Antlers
APAC         Apache                     Apache
ARD2         Ardmore                    Ardmore
ARDM         Ardmore                    Ardmore
ARNE         Arnett                     Arnett
BBOW         Broken Bow                 Broken Bow
BEAV         Beaver                     Beaver
BEEX         Bee                        Tishomingo
BESS         Bessie                     Bessie
```

Full description loaded from geoinfo.csv

A Few of our Variables… (37 in total)

Variable ID	Name	Units
2AVG	Average Wind Speed at 2m	miles per hour
2DEV	Standard Deviation of Wind Speed at 2m	miles per hour
2MAX	Maximum 2m Wind Speed	miles per hour
2MIN	Minimum 2m Wind Speed	miles per hour
9AVG	Average Air Temperature at 9m	degrees Fahrenheit
AMAX	Maximum Solar Radiation	Watts per square meter
ATOT	Total Solar Radiation	mega Joules per square meter
BAVG	Average Temperature Under Bare Soil at 10cm	degrees Fahrenheit
BMAX	Maximum Temperature Bare Soil at 10cm	degrees Fahrenheit
BMIN	Minimum Temperature Under Native Vegetation at 10cm	degrees Fahrenheit
CDEG	Cooling Degree Days	degrees Fahrenheit
DAVG	Average Dewpoint Temperature	degrees Fahrenheit
DMAX	Maximum Dewpoint Temperature	degrees Fahrenheit
DMIN	Minimum Dewpoint Temperature	degrees Fahrenheit
HAVG	Average Humidity	percent

Full description loaded from DataTranslation.csv

High-Level Task

- Load in the station and data configuration files
    - Data structures to represent each of these
- Load one or more data files
    - Each data file: many stations & days
- Compute maximum, average and minimum statistics for a given variable and station
    - These methods are for your own testing purposes – we will provide our own unit tests

Objectives

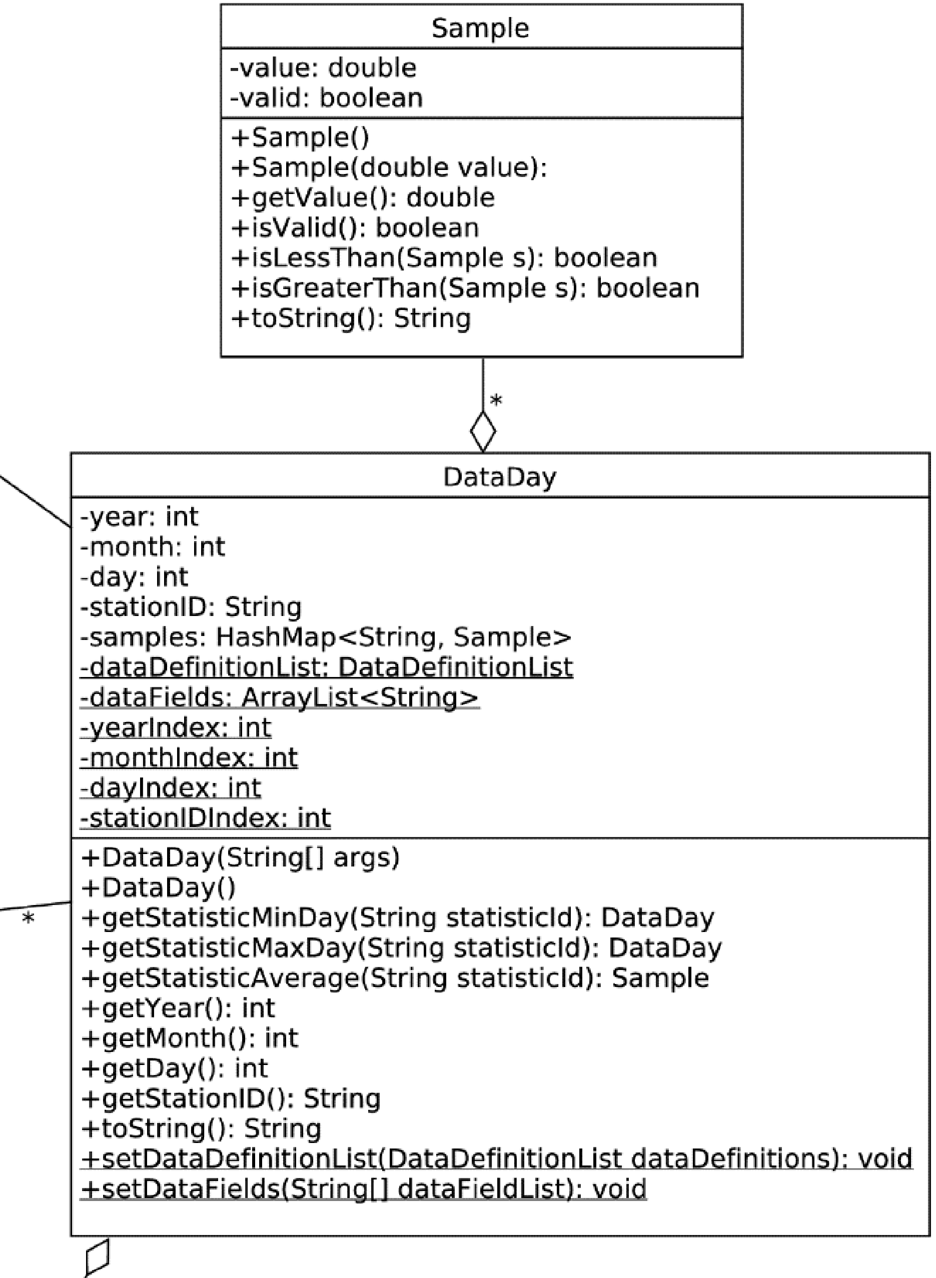
- Make use of HashMaps and TreeMaps to flexibly store data in a structure that is efficient to access
- Compute statistics over the stored data in a manner that does not rely on *a priori* knowledge of the specifics of the data
- Continue to exercise good coding practices for Javadoc and for unit testing

Code Refactor

- The structure of your classes will largely stay the same
- But: many of your existing classes will change
  - More flexibility
- And: we add a few new classes
  - Stations/data

DataDay

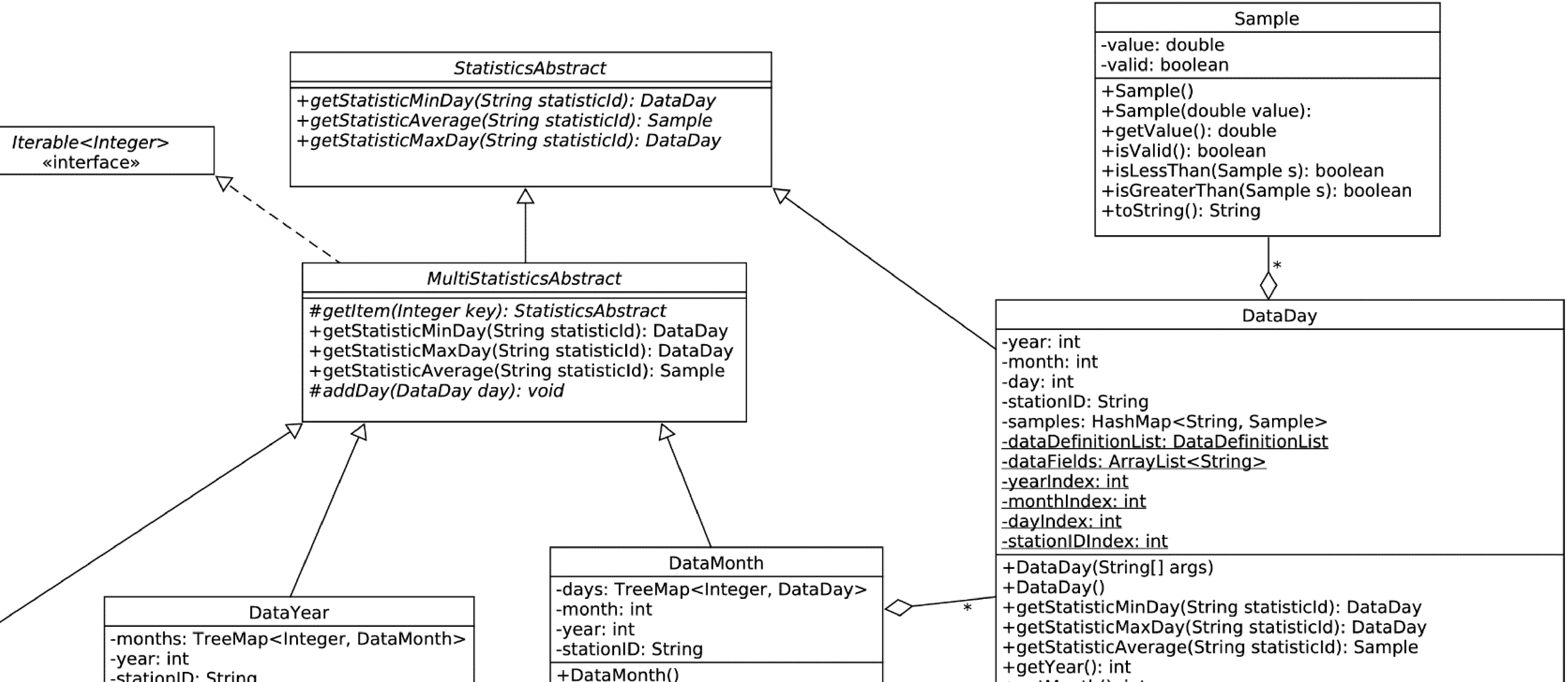
- Now store Samples in a HashMap
  - Key: stationId
- General methods for Max, Average and Min



```
Sample
-value: double
-valid: boolean
+Sample()
+Sample(double value):
+getValue(): double
+isValid(): boolean
+isLessThan(Sample s): boolean
+isGreaterThan(Sample s): boolean
+toString(): String
```

```
DataDay
-year: int
-month: int
-day: int
-stationID: String
-samples: HashMap<String, Sample>
-dataDefinitionList: DataDefinitionList
-dataFields: ArrayList<String>
-yearIndex: int
-monthIndex: int
-dayIndex: int
-stationIDIndex: int
+DataDay(String[] args)
+DataDay()
+getStatisticMinDay(String statisticId): DataDay
+getStatisticMaxDay(String statisticId): DataDay
+getStatisticAverage(String statisticId): Sample
+getYear(): int
+getMonth(): int
+getDay(): int
+getStationID(): String
+toString(): String
+setDataDefinitionList(DataDefinitionList dataDefinitions): void
+setDataFields(String[] dataFieldList): void
```

General Statistics Computation



StatisticsAbstract

+getStatisticMinDay(String statisticId): DataDay
+getStatisticAverage(String statisticId): Sample
+getStatisticMaxDay(String statisticId): DataDay

Iterable<Integer>
«interface»

MultiStatisticsAbstract

#getItem(Integer key): StatisticsAbstract
+getStatisticMinDay(String statisticId): DataDay
+getStatisticMaxDay(String statisticId): DataDay
+getStatisticAverage(String statisticId): Sample
#addDay(DataDay day): void

Sample

-value: double
-valid: boolean

+Sample()
+Sample(double value):
+getValue(): double
+isValid(): boolean
+isLessThan(Sample s): boolean
+isGreaterThan(Sample s): boolean
+toString(): String

DataDay

-year: int
-month: int
-day: int
-stationID: String
-samples: HashMap<String, Sample>
-dataDefinitionList: DataDefinitionList
-dataFields: ArrayList<String>
-yearIndex: int
-monthIndex: int
-dayIndex: int
-stationIDIndex: int

+DataDay(String[] args)
+DataDay()
+getStatisticMinDay(String statisticId): DataDay
+getStatisticMaxDay(String statisticId): DataDay
+getStatisticAverage(String statisticId): Sample
+getYear(): int

DataMonth

-days: TreeMap<Integer, DataDay>
-month: int
-year: int
-stationID: String

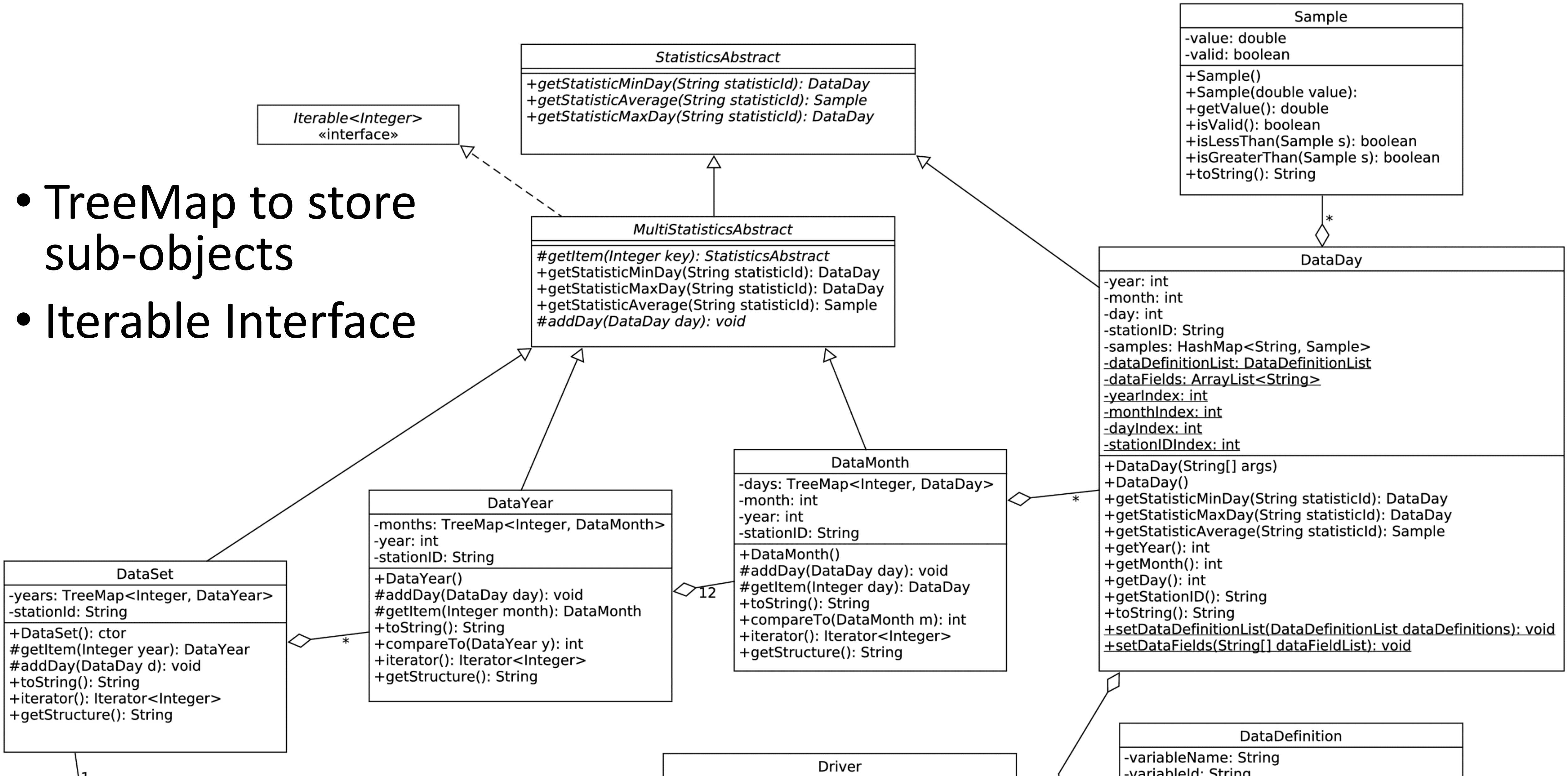
+DataMonth()

DataYear

-months: TreeMap<Integer, DataMonth>
-year: int
-stationID: String

Days to DataSets

- TreeMap to store sub-objects
- Iterable Interface



Sample
-value: double
-valid: boolean
+Sample()
+Sample(double value):
+getValue(): double
+isValid(): boolean
+isLessThan(Sample s): boolean
+isGreaterThan(Sample s): boolean
+toString(): String

StatisticsAbstract
+getStatisticMinDay(String statisticId): DataDay
+getStatisticAverage(String statisticId): Sample
+getStatisticMaxDay(String statisticId): DataDay

Iterable<Integer>
«interface»

MultiStatisticsAbstract
#getItem(Integer key): StatisticsAbstract
+getStatisticMinDay(String statisticId): DataDay
+getStatisticMaxDay(String statisticId): DataDay
+getStatisticAverage(String statisticId): Sample
#addDay(DataDay day): void

DataDay
-year: int
-month: int
-day: int
-stationID: String
-samples: HashMap<String, Sample>
-dataDefinitionList: DataDefinitionList
-dataFields: ArrayList<String>
-yearIndex: int
-monthIndex: int
-dayIndex: int
-stationIDIndex: int
+DataDay(String[] args)
+DataDay()
+getStatisticMinDay(String statisticId): DataDay
+getStatisticMaxDay(String statisticId): DataDay
+getStatisticAverage(String statisticId): Sample
+getYear(): int
+getMonth(): int
+getDay(): int
+getStationID(): String
+toString(): String
+setDataDefinitionList(DataDefinitionList dataDefinitions): void
+setDataFields(String[] dataFieldList): void

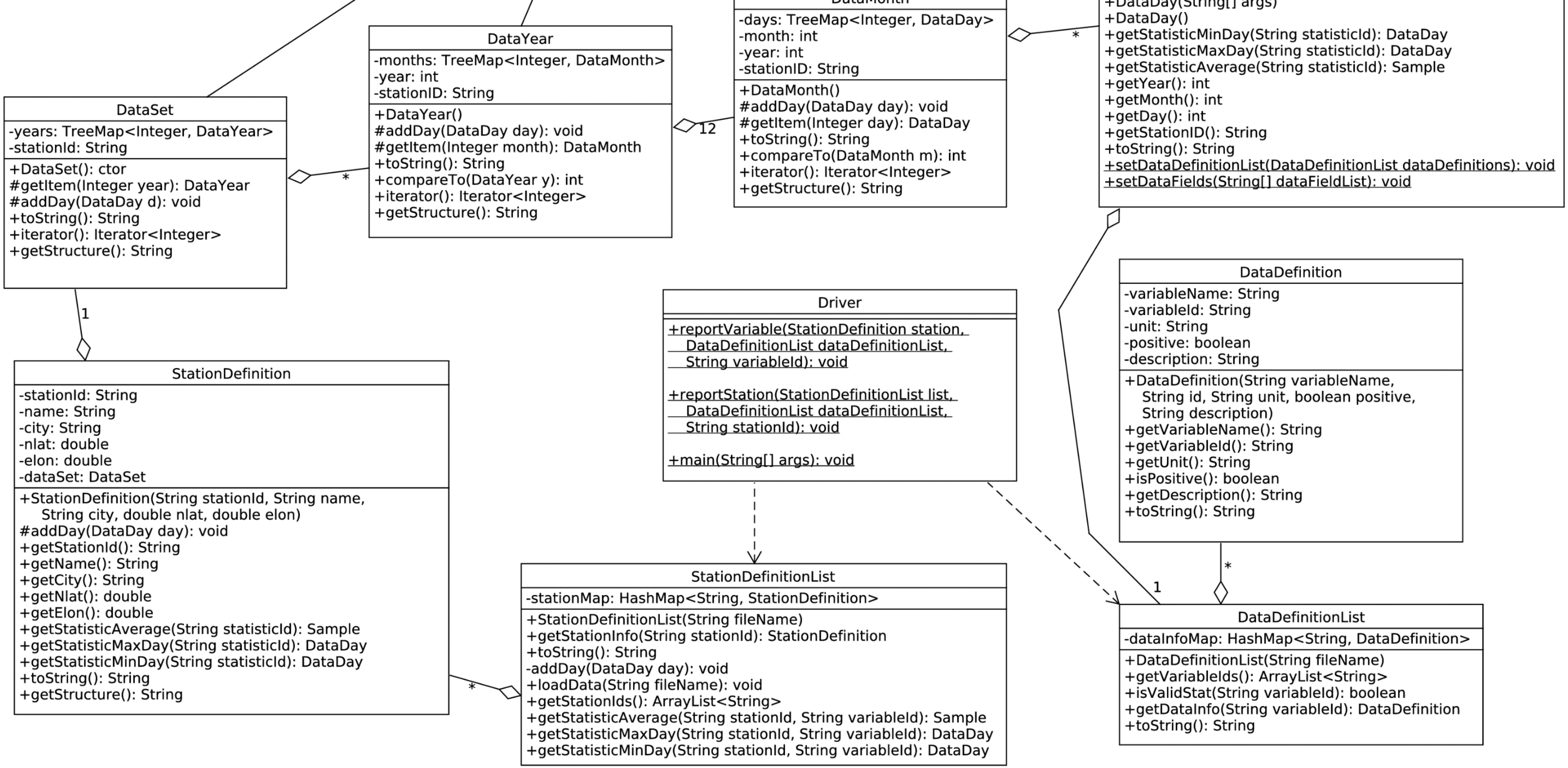
DataMonth
-days: TreeMap<Integer, DataDay>
-month: int
-year: int
-stationID: String
+DataMonth()
#addDay(DataDay day): void
#getItem(Integer day): DataDay
+toString(): String
+compareTo(DataMonth m): int
+iterator(): Iterator<Integer>
+getStructure(): String

DataYear
-months: TreeMap<Integer, DataMonth>
-year: int
-stationID: String
+DataYear()
#addDay(DataDay day): void
#getItem(Integer month): DataMonth
+toString(): String
+compareTo(DataYear y): int
+iterator(): Iterator<Integer>
+getStructure(): String

DataSet
-years: TreeMap<Integer, DataYear>
-stationId: String
+DataSet(): ctor
#getItem(Integer year): DataYear
#addDay(DataDay d): void
+toString(): String
+iterator(): Iterator<Integer>
+getStructure(): String

Driver

DataDefinition
-variableName: String
-variableId: String



DataSet

-years: TreeMap<Integer, DataYear>
-stationId: String

+DataSet(): ctor
#getItem(Integer year): DataYear
#addDay(DataDay d): void
+toString(): String
+iterator(): Iterator<Integer>
+getStructure(): String

DataYear

-months: TreeMap<Integer, DataMonth>
-year: int
-stationID: String

+DataYear()
#addDay(DataDay day): void
#getItem(Integer month): DataMonth
+toString(): String
+compareTo(DataYear y): int
+iterator(): Iterator<Integer>
+getStructure(): String

DataMonth

-days: TreeMap<Integer, DataDay>
-month: int
-year: int
-stationID: String

+DataMonth()
#addDay(DataDay day): void
#getItem(Integer day): DataDay
+toString(): String
+compareTo(DataMonth m): int
+iterator(): Iterator<Integer>
+getStructure(): String

+DataDay(String[] args)
+DataDay()
+getStatisticMinDay(String statisticId): DataDay
+getStatisticMaxDay(String statisticId): DataDay
+getStatisticAverage(String statisticId): Sample
+getYear(): int
+getMonth(): int
+getDay(): int
+getStationID(): String
+toString(): String
+setDataDefinitionList(DataDefinitionList dataDefinitions): void
+setDataFields(String[] dataFieldList): void

DataDefinition

-variableName: String
-variableId: String
-unit: String
-positive: boolean
-description: String

+DataDefinition(String variableName,
  String id, String unit, boolean positive,
  String description)
+getVariableName(): String
+getVariableId(): String
+getUnit(): String
+isPositive(): boolean
+getDescription(): String
+toString(): String

Driver

+reportVariable(StationDefinition station,
  DataDefinitionList dataDefinitionList,
  String variableId): void

+reportStation(StationDefinitionList list,
  DataDefinitionList dataDefinitionList,
  String stationId): void

+main(String[] args): void

StationDefinition

-stationId: String
-name: String
-city: String
-nlat: double
-elon: double
-dataSet: DataSet

+StationDefinition(String stationId, String name,
  String city, double nlat, double elon)
#addDay(DataDay day): void
+getStationId(): String
+getName(): String
+getCity(): String
+getNlat(): double
+getElon(): double
+getStatisticAverage(String statisticId): Sample
+getStatisticMaxDay(String statisticId): DataDay
+getStatisticMinDay(String statisticId): DataDay
+toString(): String
+getStructure(): String

StationDefinitionList

-stationMap: HashMap<String, StationDefinition>

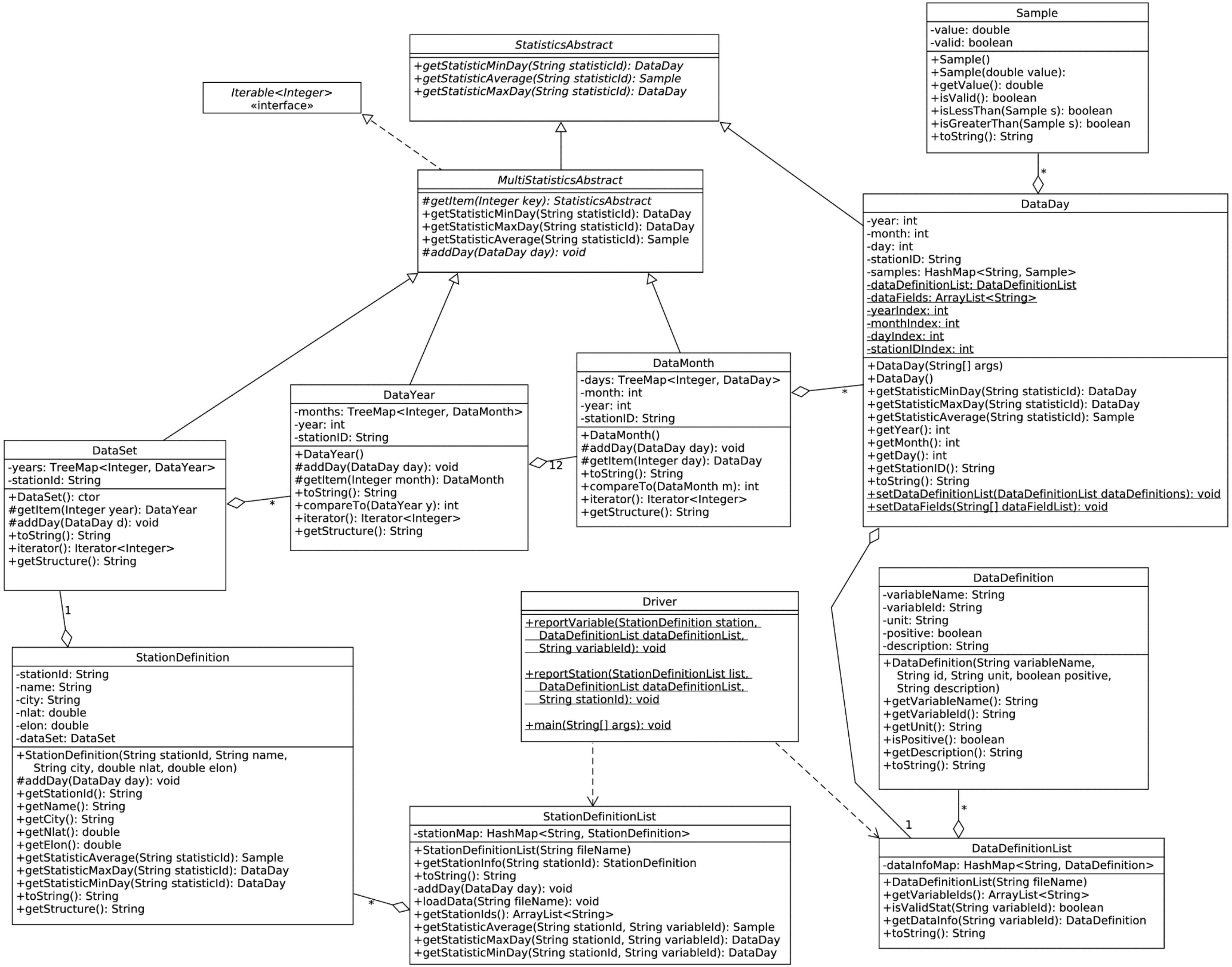
+StationDefinitionList(String fileName)
+getStationInfo(String stationId): StationDefinition
+toString(): String
-addDay(DataDay day): void
+loadData(String fileName): void
+getStationIds(): ArrayList<String>
+getStatisticAverage(String stationId, String variableId): Sample
+getStatisticMaxDay(String stationId, String variableId): DataDay
+getStatisticMinDay(String stationId, String variableId): DataDay

DataDefinitionList

-dataInfoMap: HashMap<String, DataDefinition>

+DataDefinitionList(String fileName)
+getVariableIds(): ArrayList<String>
+isValidStat(String variableId): boolean
+getDataInfo(String variableId): DataDefinition
+toString(): String

New Classes



Iterable<Integer>
«interface»

StatisticsAbstract

+getStatisticMinDay(String statisticId): DataDay
+getStatisticAverage(String statisticId): Sample
+getStatisticMaxDay(String statisticId): DataDay

Sample

-value: double
-valid: boolean

+Sample()
+Sample(double value):
+getValue(): double
+isValid(): boolean
+isLessThan(Sample s): boolean
+isGreaterThan(Sample s): boolean
+toString(): String

MultiStatisticsAbstract

#getItem(Integer key): StatisticsAbstract
+getStatisticMinDay(String statisticId): DataDay
+getStatisticMaxDay(String statisticId): DataDay
+getStatisticAverage(String statisticId): Sample
#addDay(DataDay day): void

DataDay

-year: int
-month: int
-day: int
-stationID: String
-samples: HashMap<String, Sample>
-dataDefinitionList: DataDefinitionList
-dataFields: ArrayList<String>
-yearIndex: int
-monthIndex: int
-dayIndex: int
-stationIDIndex: int

+DataDay(String[] args)
+DataDay()
+getStatisticMinDay(String statisticId): DataDay
+getStatisticMaxDay(String statisticId): DataDay
+getStatisticAverage(String statisticId): Sample
+getYear(): int
+getMonth(): int
+getDay(): int
+getStationID(): String
+toString(): String
+setDataDefinitionList(DataDefinitionList dataDefinitions): void
+setDataFields(String[] dataFieldList): void

DataMonth

-days: TreeMap<Integer, DataDay>
-month: int
-year: int
-stationID: String

+DataMonth()
#addDay(DataDay day): void
#getItem(Integer day): DataDay
+toString(): String
+compareTo(DataMonth m): int
+iterator(): Iterator<Integer>
+getStructure(): String

DataYear

-months: TreeMap<Integer, DataMonth>
-year: int
-stationID: String

+DataYear()
#addDay(DataDay day): void
#getItem(Integer month): DataMonth
+toString(): String
+compareTo(DataYear y): int
+iterator(): Iterator<Integer>
+getStructure(): String

DataSet

-years: TreeMap<Integer, DataYear>
-stationId: String

+DataSet(): ctor
#getItem(Integer year): DataYear
#addDay(DataDay d): void
+toString(): String
+iterator(): Iterator<Integer>
+getStructure(): String

DataDefinition

-variableName: String
-variableId: String
-unit: String
-positive: boolean
-description: String

+DataDefinition(String variableName,
String id, String unit, boolean positive,
String description)
+getVariableName(): String
+getVariableId(): String
+getUnit(): String
+isPositive(): boolean
+getDescription(): String
+toString(): String

Driver

+reportVariable(StationDefinition station,
DataDefinitionList dataDefinitionList,
String variableId): void

+reportStation(StationDefinitionList list,
DataDefinitionList dataDefinitionList,
String stationId): void

+main(String[] args): void

StationDefinition

-stationId: String
-name: String
-city: String
-nlat: double
-elon: double
-dataSet: DataSet

+StationDefinition(String stationId, String name,
String city, double nlat, double elon)
#addDay(DataDay day): void
+getStationId(): String
+getName(): String
+getCity(): String
+getNlat(): double
+getElon(): double
+getStatisticAverage(String statisticId): Sample
+getStatisticMaxDay(String statisticId): DataDay
+getStatisticMinDay(String statisticId): DataDay
+toString(): String
+getStructure(): String

StationDefinitionList

-stationMap: HashMap<String, StationDefinition>

+StationDefinitionList(String fileName)
+getStationInfo(String stationId): StationDefinition
+toString(): String
-addDay(DataDay day): void
+loadData(String fileName): void
+getStationIds(): ArrayList<String>
+getStatisticAverage(String stationId, String variableId): Sample
+getStatisticMaxDay(String stationId, String variableId): DataDay
+getStatisticMinDay(String stationId, String variableId): DataDay

DataDefinitionList

-dataInfoMap: HashMap<String, DataDefinition>

+DataDefinitionList(String fileName)
+getVariableIds(): ArrayList<String>
+isValidStat(String variableId): boolean
+getDataInfo(String variableId): DataDefinition
+toString(): String

Notes

- Data loading is managed by StationDefinitionList
  - All subordinate classes provide an addDay() method that places the day into the correct location in the data structure
- getStatisticMin(), Max(), Average(): variable type is encoded in the String variableId and not in the method name
- We have specified a number of String-returning methods in the UML diagram. These are most often useful for debugging (we won't be testing their output). More details in the specification

Deadlines

- Project must be submitted by Wednesday, Oct 26th @1:29pm
- Code review must be completed by Wednesday, Nov 2nd