

Projektplan – Verbrechen in Baltimore – Philipp Nowak

Was soll umgesetzt werden?

Ich möchte mit diesem Projekt untersuchen, wie es um die Sicherheit in Baltimore steht. Als Grundlage dient eine CSV-Datei, die von der US-Regierung auf <http://catalog.data.gov/dataset/crime-safety-2010-2013> veröffentlicht wurde.

Folgende Fragen sollen beantwortet werden: Hat die Verbrechenslage im Laufe der letzten Jahre zu- oder abgenommen? In welchen Stadtbezirken lebt es sich am sichersten, in welchen am gefährlichsten? Gibt es eine Korrelation zwischen einzelnen Verbrechenarten und der Verbrechenslage? Wie ist die Entwicklung im „sichersten“ und „gefährlichsten“ Bezirk hinsichtlich mehrerer Faktoren (Verbrechensrate, Gewalt, Häusliche Gewalt, Schusswaffentötung)? Und wie sieht die Entwicklung bezüglich der Faktoren in ganz Baltimore über die Jahre hinweg aus?

Diese Fragen sollen mithilfe von selbst erstellten Diagrammen sowie deren Deutung beantwortet werden. Als Bonus obendrauf ist eine eigene Recherche über die auffälligsten Stadtbezirke geplant, wo geschaut wird, ob meine Ergebnisse zu begründen sind oder ob es zu Überraschungen kommt.

Wie soll es umgesetzt werden?

Umgesetzt werden soll mein Projekt mit Python.

Der Aufbau meiner Dokumentation zur Beantwortung der Eingangsfragen soll wie folgt aussehen:

Einleitung:

- Durchschnitt Verbrechenquotient der Jahre 2010-2014 in Baltimore absolut (x-Achse: Jahre)
- Median Verbrechenquotient der Jahre 2010-2014 (x-Achse: Jahre)
- Modalwert Verbrechenquotient der Jahre 2010-2014 (x-Achse: Jahre)
- Varianz Verbrechenquotient der Jahre 2010-2014 (x-Achse: Jahre)
- Standardabweichung Verbrechenquotient der Jahre 2010-2014 (x-Achse: Jahre)
- Spannweite Verbrechenquotient der Jahre 2010-2014 (x-Achse: Jahre)
- Bei wie viel Prozent der Bezirke liegt die Kriminalitätsrate unter dem Wert „50“ bei den Jahren 2010-2014? (x-Achse: Jahre)

Untersuchung Verbrechen (crime):

- Top 10 Verbrechen in 2010 (x-Achse: Bezirke)
- Flop 10 Verbrechen in 2010 (x-Achse: Bezirke)
- Top 10 Verbrechen in 2012 (x-Achse: Bezirke)
- Flop 10 Verbrechen in 2012 (x-Achse: Bezirke)
- Top 10 Verbrechen in 2014 (x-Achse: Bezirke)
- Flop 10 Verbrechen in 2014 (x-Achse: Bezirke)
- Veränderung der Top-Verbrechen über die Jahre (x-Achse: Jahre)
- Veränderung der Flop-Verbrechen über die Jahre (x-Achse: Jahre)
- Wie oft sind die Verbrechenswerte (übergreifend) über die Jahre gestiegen im Vergleich zu vorher? (x-Achse: Bezirke)

- Wie oft sind die Verbrechenswerte (übergreifend) über die Jahre gefallen im Vergleich zu vorher? (x-Achse: Bezirke)
- Prozentuale Veränderung des Verbrechensquotients aus dem Jahre 2010 im Vergleich zum Jahre 2014 bei den Bezirken der Top10-Liste (x-Achse: Bezirke)
- Prozentuale Veränderung des Verbrechensquotients aus dem Jahre 2010 im Vergleich zum Jahre 2014 bei den Bezirken der Flop10-Liste (x-Achse: Bezirke)

Vergleich zu Unterkategorie 1 von Verbrechen: Gewalt (viol):

- Top 10 viol in 2014 (x: Areale)
- Flop 10 viol in 2014 (x: Areale)
- Veränderung der Top-viol über die Jahre (x: Jahre)
- Veränderung der Flop-viol über die Jahre (x: Jahre)

Vergleich zu Unterkategorie 2 von Verbrechen: Diebstahl (prop):

- Top 10 prop in 2014 (x: Areale)
- Flop 10 prop in 2014 (x: Areale)
- Veränderung der Top-prop über die Jahre (x: Jahre)
- Veränderung der Flop-prop über die Jahre (x: Jahre)

Abschlussuntersuchung:

- Liniendiagramm „gefährlichstes“ Viertel (anhand Durchschnittswert aller verfügbaren Daten) über Jahre: crime, viol, prop, domvio (Häusliche Gewalt), gunhom (Schusswaffentötung): Wie hat es sich verändert, gibt es Auffälligkeiten?
- Liniendiagramm „sicherstes“ Viertel (anhand Durchschnittswert aller verfügbaren Daten) über Jahre: crime, viol, domvio, gunhom: Wie hat es sich verändert, gibt es Auffälligkeiten?
- Liniendiagramm Baltimore über Jahre: Durchschnittswerte crime, viol, prop, domvio, gunhom: Was hat sich wie stark verändert?

Folgende Funktionen wird das Programm enthalten:

- Funktion zur Berechnung der Summe einer Liste
- Funktion zur Berechnung des Durchschnitts (Funktion „Summe einer Liste“ und Länge der Liste (len) benötigt)
- Funktion zur Berechnung der „Top 10“ (Die höchsten zehn Werte inklusive jeweils zugehöriger Stadtbezirke; als Liste)
- Funktion zur Berechnung der „Flop 10“ (Die niedrigsten zehn Werte inklusive jeweils zugehöriger Stadtbezirke; als Liste)
- Funktion zur Berechnung des Medians
- Funktion Funktion zur Berechnung, wie oft der Wert im Vergleich zu vorher gestiegen ist (= in welchen Bezirken wird die Kriminalitätsrate kontinuierlich schlechter?)
- Funktion Funktion zur Berechnung, wie oft der Wert im Vergleich zu vorher gefallen ist (= in welchen Bezirken verbessert sich die Kriminalitätsrate kontinuierlich?)
- Funktion zur Berechnung des Modalwerts
- Funktion zur Berechnung der Varianz
- Funktion zur Berechnung der Standardabweichung
- Funktion zur Berechnung der Spannweite

- Funktion zur Berechnung des Prozentsatzes, bei wie vielen Bezirken sich die Kriminalitätsrate unter dem Wert „50“ befindet
- Funktion zur Berechnung des Prozentsatzes, wie sich der Verbrechensquotient im Vergleich vom Jahre 2010 im Vergleich zum Jahre 2014 verändert hat

Was dafür getan werden muss:

- Öffnen und einlesen der CSV-Datei (import csv benötigt)
- Listen definieren, mit denen ich arbeiten möchte
- Listen füllen anhand der CSV-Datei
- Störende Spaltenüberschriften entfernen für die Listen
- Umwandeln der Strings in Integers, damit mit ihnen gearbeitet werden kann
- Die vier oben genannten Funktionen werden für die Berechnungen gebraucht
- Berechnung Verbrechens-Durchschnittswert pro Jahr
- Berechnung der jeweiligen Top-10
- Berechnung der jeweiligen Flop-10
- Berechnung der jeweiligen Top-1 für Veränderung der Top-Verbrechens/Gewalt/Diebstahl-Werte über die Jahre hinweg
- Berechnung der jeweiligen Flop-1 für Veränderung der Top-Verbrechens/Gewalt/Diebstahl-Werte über die Jahre hinweg
- Berechnung des Durchschnitts für alle Werte je Stadtbezirk für Abschlussdiagramm
- Die Berechnung der weiteren Funktionen

- Für die Darstellung von Diagrammen bin ich letztes Semester im Projektseminar Informetrie mit matplotlib in Berührung gekommen, für die Darstellung der Diagramme in diesem Projekt würde ich mich in ein anderes System reinarbeiten (Pandas? Seaborn?)
- Die jeweiligen Verbrechensarten sollen ein eigenes Farbthema kriegen; verschiedene Jahre mit einer Verbrechensart sollen ebenfalls farblich unterschieden werden (heller/dunkler)
- Als Diagramm-Arten sind Balkendiagramme und für den Dreier-Abschluss Flussdiagramme, wo die verschiedenen Verbrechensarten farblich unterscheidbar sind, geplant
- Um einen erhöhten Arbeitsaufwand zu gewährleisten, ist die Einbettung meiner Arbeit in eine GUI vorgesehen (Tkinter?)
- Optionale Erweiterung: Verschiedene Systeme der Diagramm-Darstellung (matplotlib, Pandas, Seaborn ?) in der GUI anzeigen lassen
- Optionale Erweiterung: Noch mehr Varianz von Diagramm-Arten, sofern es Sinn macht, nutzen