Sind deutsche Autos anders als ausländische?

HELMUT WIRTHS, OLDENBURG

Zusammenfassung: *Vorgestellt werden Überlegungen zur Vorbereitung einer Unterrichtsreihe zur Statistik, in der Methoden und Begriffe der explorativen Datenanalyse EDA benutzt werden, ebenso Arbeitsergebnisse aus dem Unterricht sowie Beobachtungen beim Umgang mit den Begriffen und Methoden der EDA.*

1 Einführung

Wer kennt nicht Pauschalurteile über deutsche und ausländische Autos. Französische PKW seien beispielsweise futuristischer, italienische eleganter und sportlicher und amerikanische Straßenkreuzer grösser und geräumiger als deutsche, mag man von der einen oder anderen Seite hören. Solche oder ähnliche Äußerungen dürften jedem von uns schon begegnet sein. Über Geschmack (modern, futuristisch, elegant) mag man streiten, diesbezügliche Meinungen könnte man über Umfragen ermitteln. Ich möchte diesen Aspekt, so lohnend seine Verfolgung und so wichtig eine Erziehung zur seriösen Erforschung von Meinungen auch ist, hier ausblenden. In diesem Beitrag soll verdeutlicht werden, dass sich eine Unterrichtseinheit zur EDA lohnt.

2 Zur Datenerhebung

In einschlägigen Zeitschriften, Zeitungen, auf CD-Rom oder im Internet findet man Informationen über die auf dem Markt vorhandenen Pkw-Modelle. Informationslücken wurden vorwiegend mit Hilfe der „ADAC Special Auto CD-Rom“ geschlossen. Beim letzten Unterrichtsdurchgang war es die CD „Auto ‘98“. Nach eingehender Diskussion des von den Lernenden mitgebrachten Materials hat sich die Lerngruppe darauf geeinigt, folgende sechs Merkmale bei jedem Autotyp mit Ottomotor zu erfassen:

- Hubraum in cm^3
- Leistung in kW
- Gewicht in kg
- Verbrauch in Liter pro 100 km
- Höchstgeschwindigkeit in Kilometer pro Stunde
- Zeit für die Beschleunigung von 0 auf 100 km/h

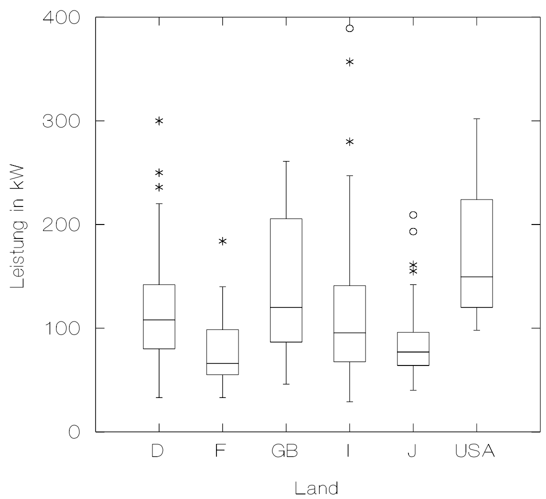
Da die Leistungsangabe in PS mit einem festen Umrechnungsfaktor in kW umgerechnet werden kann, haben wir trotz der noch erstaunlich starken Verankerung der „Pferdestärken“ in den Köpfen auf die Erfassung der Leistung in PS verzichtet. Ursprünglich wollten die Lernenden auch Zeiten für die Beschleunigung von zum Beispiel 60 auf 120 km/h erfassen. Doch zeigte sich bald, dass die entsprechenden Zahlen nicht immer oder nur mit großem Aufwand zu beschaffen waren. Wir hofften aber, mit diesen sechs Merkmalen insgesamt genügend Informationen für die geplanten Untersuchungen bereitstellen zu können. Als Länder wurden ausgewählt: Deutschland (D), Frankreich (F), Großbritannien (GB), Italien (I), Japan (J) und USA. Aus Russland, Schweden, Spanien, der Tschechoslowakei und aus Korea konnten nur wenige vollständige Datensätze zusammengestellt werden. Die Lernenden haben sich daher entschlossen, Datensätze dieser Länder nicht mit einzubeziehen, zumal sie meinten, mit insgesamt 150 Datensätzen aus den sechs ausgewählten Ländern genügend Untersuchungsmaterial zu besitzen.

Zur Auswertung konnte jeder Lernende die Statistikmöglichkeiten seines graphikfähigen Taschenrechners TI-82 einsetzen. Außerdem wurden von den Lernenden an den Schulcomputern die Tabellenkalkulation Microsoft Excel sowie auf meinem eigenen Rechner das Statistikprogramm Student Systat 1.0 for Windows genutzt.

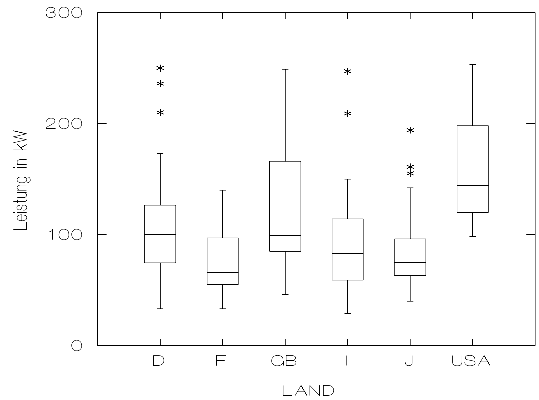
3 Zu den Ergebnissen

In den 8. Klassen habe ich mit Datensätzen kleinen Umfangs die einfachen Boxplots erstellen lassen, bei denen die Whiskers vom Minimum bis zum 1. Quartil sowie vom 3. Quartil zum Maximum gezeichnet werden. Dies ist für mich der „Boxplot der beschreibenden Statistik“. Mit graphikfähigen Taschenrechnern vom TI-82 an aufwärts kann man diesen Typ erstellen. In Grundkursen der gymnasialen Oberstufe habe ich dem Wunsch der Schüler entsprochen und möglichst viele Daten aus den sechs Ländern sammeln lassen. Zuletzt wurde diese Datensammlung auf insgesamt 150 Datensätze erweitert (davon 43 aus D, 15 aus F, 19 aus GB, 24 aus I, 39 aus J und 10 aus USA). Schon bei der Daten-Erfassung gab es genug Stoff für Diskussionen darüber, ob ein Automodell in die Datensammlung aufgenommen werden soll. Die Lernenden hatten sich geeinigt, in Grenzfällen die Daten der Pkw zunächst einmal aufzunehmen und nach Erstellung der Boxplots über ihre Beibehaltung in der Datensammlung zu diskutieren und zu entscheiden. Wesentliche Grundlage für diese Entscheidung war der zweite Typ für Boxplots, den ich „Boxplot der beurteilenden Statistik“ nenne, eine Option, die ich beim TI-83, TI-89 und TI-92 Plus habe. Die in diesem Aufsatz abgebildeten Boxplots des zweiten Typs wurden mit Hilfe von Student Systat erstellt. Die Whiskers werden nur vom kleinsten Datenwert größer oder gleich 1. Quartil - 1,5R bis zum 1. Quartil sowie vom 3. Quartil bis zum größten Datenwert kleiner oder gleich 3. Quartil + 1,5R gezeichnet, wobei der Interquartilsabstand R als Abstand von 3. Quartil und 1. Quartil definiert ist. Mit „*" werden alle die Daten in den Boxplot eingezeichnet, die entweder mehr als 1,5R, aber nicht mehr als 3R unterhalb vom 1. Quartil oder aber mehr als 1,5R, aber nicht mehr als 3R oberhalb des 3. Quartils liegen. Alle Daten, die weiter außerhalb liegen, also entweder mehr als 3R unterhalb des 1. Quartils oder um mehr als 3R oberhalb des 3. Quartils, werden mit „o" eingezeichnet. Weitere Hinweise hierzu in Abschnitt 4.

Zur Leistung:

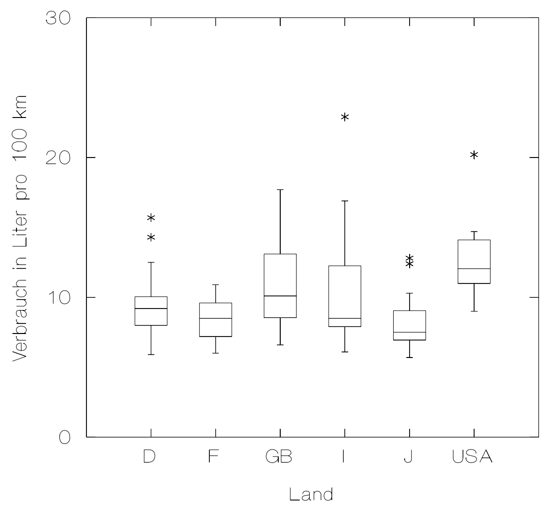


Bei Boxplots der beschreibenden Statistik sollte man einen Bereich, der durch extrem lange Whiskers dargestellt wird, näher untersuchen. Beim oben dargestellten Boxplottyp erregten zunächst alle Daten die größte Aufmerksamkeit, die durch „*" oder „o" eingezeichnet worden sind. So werden mögliche Ausreißer gekennzeichnet. Die Lernenden sprachen jedoch von besonderen Automodellen, die bewusst für einen bestimmten Zweck produziert werden, und daher keine rein zufälligen Ausreißer darstellen. Im obigen Boxplot verbergen sich hinter den Zeichen „*" oder „o" zum einen Autos, die eher Sportwagen als normale Gebrauchs-Pkw sind, zum anderen aber Luxuslimousinen. Es sind dies im einzelnen: 2 Porsche, 1 Audi, 1 Renault, 2 Ferrari, 1 Lamborghini, 1 Mitsubishi, 2 Lexus, 1 Mazda. Die meisten dieser Autos haben bereits bei der Datensammlung Aufmerksamkeit erregt und wurden von der Lerngruppe nur mit Bedenken vorläufig in die Datensammlung aufgenommen. Nach Vergleich der Boxplots für die Leistung entschied die Lerngruppe: Alle Sportwagen werden aus der Datensammlung herausgenommen, die Luxuslimousinen werden jedoch beibehalten. Das bedeutete, dass noch sechs weitere Modelle (1 Porsche, 1 Chrysler, 2 Aston Martin und 2 Lotus) entfernt werden mussten. Jetzt aber zögerten die Lernenden. Nun fehle das „Salz in der Suppe", die Boxplots würden langweiliger, es gäbe dann zwar weiterhin noch genug Anlässe für Diskussionen, Unterschiede seien jetzt aber mühsamer zu entdecken und zu begründen, die motivierenden Aufhänger für Gespräche fehlten. Soweit die Lernenden. Und hier der neue Boxplot auf der Basis der restlichen 137 Datensätze *ohne die Sportwagen:*



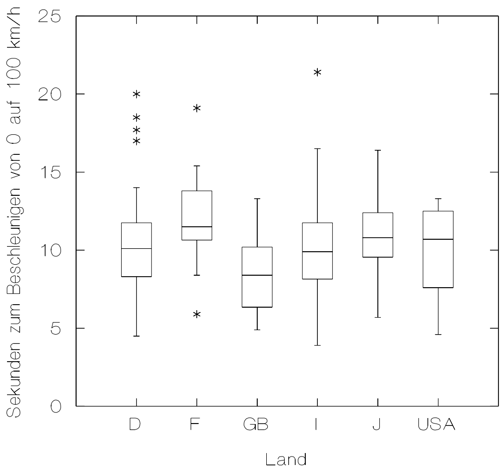
Die unterschiedlichen Kennzahlen (1., 3. Quartil, Median), die Länge der Whiskers, der Interquartilsabstand, die besonders mit „*" oder „o" gekennzeichneten Modelle sowie die Änderungen gegenüber dem ersten Boxplot, all das reizt zu Diskussionen und zum Vergleich. Die Boxplots sprechen für sich. Die übrigen Boxplots werde ich auf der Basis von allen 150 Datensätzen vorstellen.

Zum Verbrauch:



Wir unterstellen einmal, dass man mit Hilfe der oben gegebenen Boxplots entscheiden kann, in welchem Land Autobauer und Kunden am meisten auf möglichst geringen Verbrauch achten. Nach Augenmaß kann man sofort Großbritannien, Italien und USA ausschließen. Im Vergleich mit Frankreich scheidet Deutschland aus. Gegen Deutschland sprechen der größere Median, das größere 1. und 3. Quartil sowie 2 Ausreißer mit hohem Verbrauch. Wir entscheiden uns für Japan im Vergleich mit Frankreich. Für Autos aus Japan sprechen der kleinere Median und das kleinere 1. und 3. Quartil, dagegen drei Ausreißer mit höherem Verbrauch. Einer davon ist aber als Sportwagen einzustufen. Statt kleineres 1. und 3. Quartil wurde auch mit kleinerem 1. Quartil und geringerem Interquartilsabstand argumentiert. Auch die Benzin„säufer" erregten großes Interesse. Bei Japan waren es zwei Luxuslimousinen, in allen anderen Fällen aber Sportwagen, die sich mit besonders hohem Verbrauch so von den anderen Autos abheben, dass sie im Boxplot mit „*" gekennzeichnet werden müssen.

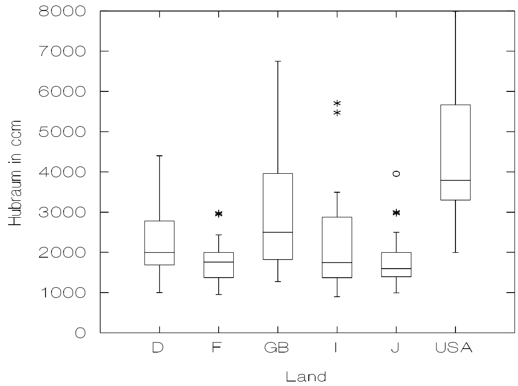
Zur Beschleunigung:



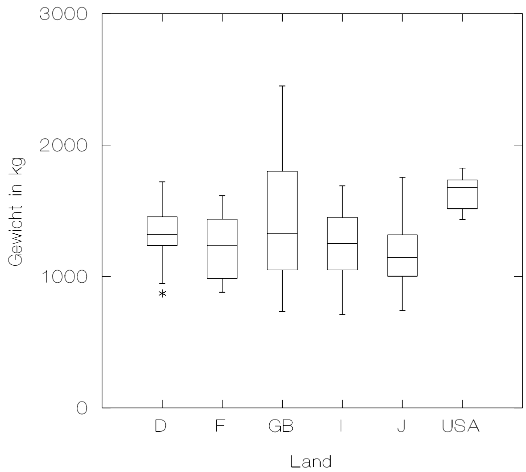
Wir können anhand dieser Boxplots nicht belegen, dass Italien Musterland für sportliche Autos ist. Insofern muss ein früher im Unterricht genanntes Vorurteil revidiert werden. Das Auto mit der besten Beschleunigung stammt zwar aus Italien, aber auch das lahmste Fahrzeug. Beseitigt man alle Sportwagen, dann sieht die Situation für Italien in Bezug auf Spitzenbeschleunigungen noch schlechter aus. Wenn man ein Land mit den insgesamt sportlichsten Autos nennen möchte, sollte nach Meinung der Lerngruppe die Wahl auf Großbritannien fallen.

Zum Hubraum:

Hier werden alle Meinungen über die extrem großvolumigen Pkw aus den USA bestätigt, denen Pkw aus Großbritannien und die japanischen Luxusautos am nächsten kommen. Diese Tendenz wird noch deutlicher, wenn man die Grenze zwischen Sportwagen und normalen Gebrauchs-Pkw wie oben erwähnt so zieht, dass Sportwagen eliminiert werden.

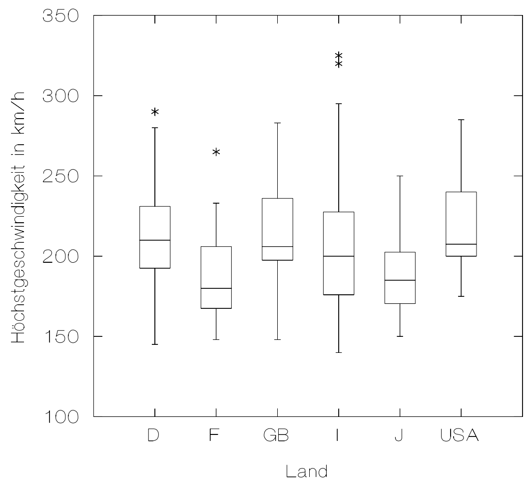


Zum Gewicht:



Hier überrascht die enorme Bandbreite bei den Wagen aus Großbritannien vom Rolls Royce bis zum Rover Mini, weniger die besonders exponierte Stellung der Autos aus den USA. Der unterhalb der unteren Whiskers eingezeichnete Ford Ka ist eine Ausnahme in Deutschland, wird aber von Autos anderer Länder noch unterboten.

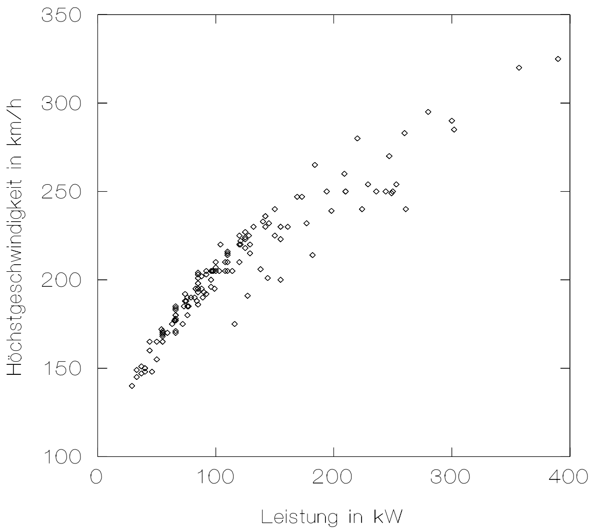
Zur Höchstgeschwindigkeit:



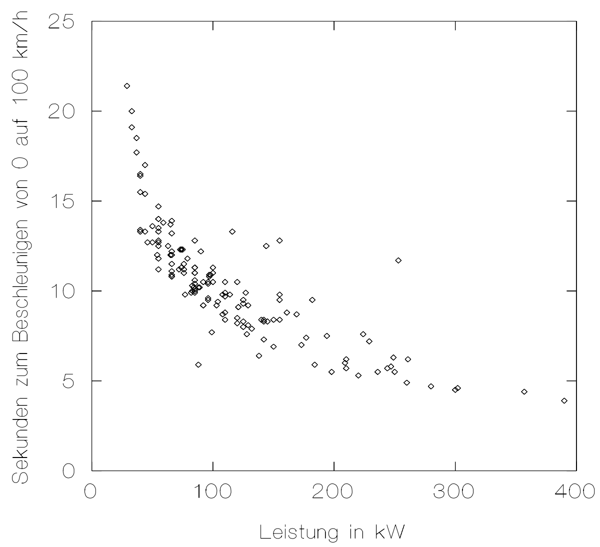
Nach den bisherigen Boxplots sollte man zum Teil erhebliche Unterschiede erwarten. Das dies nicht der Fall ist, erstaunt. Diese Tendenz wird noch deutlicher, wenn man die Grenze zwischen Sportwagen und normalen Gebrauchs-Pkw wie oben erwähnt so zieht, dass die Sportwagen eliminiert werden.

Zum Abschluss werden noch drei Graphen zu allgemein vorhandenen (Vor-)Urteilen vorgestellt. Diese Graphen wurden auf der Basis aller 150 Datensätze erstellt. Nun wird nicht mehr nach Ländern unterschieden.

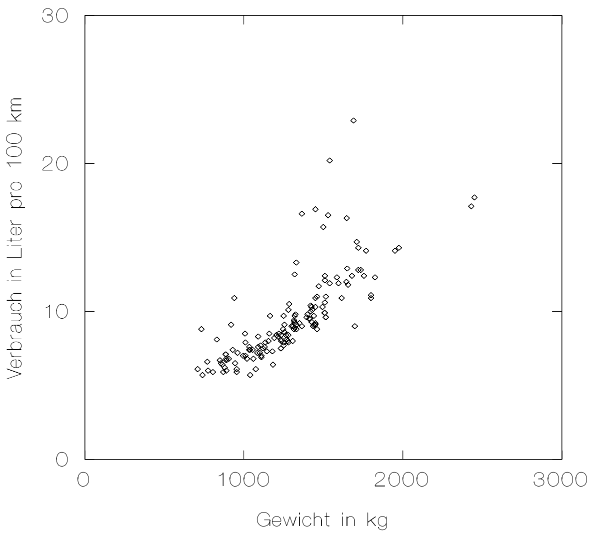
1. Je leistungsstärker, desto schneller.



2. Je leistungsstärker, desto rasanter.



3. Je leichter, desto sparsamer.



Kann man sich der Suggestion dieser drei Graphiken entziehen? Wer möchte nicht eine Modellierung versuchen, eine Ausgleichskurve durch jede dieser drei Punktwolken legen und Prognosewerte berechnen, aber auch „Ausreißer“ bestimmen? Dass diese Zuordnungen nur Relationen und keine Funktionen darstellen, wurde den Lernenden bereits während der Datensammlung klar. Als ein Beispiel unter vielen sei die Zuordnung Hubraum → Leistung genannt. Um bei diesen drei Punktwolken Mehrdeutigkeiten stärker auszuschließen, diskutierten die Lernenden weitere Einschränkungen hinsichtlich der Datenauswahl wie zum Beispiel Beschränkung auf 4-Zylinder-Motoren, bei mehreren Motorvarianten eine Beschränkung auf einen Motor, der maximale Leistung im mittleren Drehzahlbereich erreicht, oder Verzicht auf Motoren in Mehrventiltechnik. Die Lernenden erhofften sich mit solchen zusätzlichen Einschränkungen, dass bei diesen zweidimensionalen Dastellungen nun mögliche Trends deutlicher werden.

4 Zum Unterricht

In einer 8. Klasse haben wir zunächst mit kleinen Datensätzen von 6 bis 12 Elementen gearbeitet. Für jedes Element des Datensatzes wurde eine eigene Karteikarte angefertigt. Elemente eines Datensatzes waren die einzelnen Automodelle. Sortiert wurde der Karteikartenstapel nach dem uns jeweils interessierenden Merkmal. Die Bestimmung von Maximum und Minimum eines bestimmten Merkmals stellte bei einem korrekt sortierten Karteikartenstapel kein Problem dar. Interessant war für mich die Diskussion zum Median. Zunächst wollten die Schülerinnen und Schüler die Kennzahlen - im ersten Teil ging es nur um Maximum, Minimum und Median, die Behandlung der beiden Quartile erfolgte später - durch farbige Reiter auf der zugehörigen Karteikarte markieren. Für Maximum und Minimum war das immer möglich, beim Median jedoch nur bei ungerader Zahl an Karteikarten. Bei gerader Kartenzahl zerfiel der Stapel in zwei gleich große Teile und es gab keine Karte „genau in der Mitte“, auf der der Median markiert werden konnte. Als Ausweg nahmen die Schülerinnen und Schüler einfach eine noch nicht beschriebene Karteikarte, stellten sie quer zum Stapel, um anzudeuten, dass es sich nicht um eine Datenkarte handelte, und notierten schließlich den Median auf dieser besonderen Karte. Von einer Markierung durch farbige Reiter war danach nicht mehr die Rede. Bei ungerader Kartenzahl kamen sie auf folgende Idee: Jede Karte wird ein zweites Mal hergestellt. Nach dieser Verdopplung liegt ein Stapel mit gerader Anzahl Karten vor. Jetzt kann dieser Kartenstapel in zwei gleich große Teile geteilt und der Median auf einer besonderen Karte markiert werden. Zu ihrem Vorgehen merkten die Lernenden noch an, dass Maximum, Median, Minimum und auch der arithmetische Mittelwert von dieser Kartenverdopplung nicht verändert werden, dass sie sich nun aber die Vorgänge besser vorstellen können. Dass der Median bei gerader Anzahl an Karteikarten als Mittelwert der Daten der letzten Karte des ersten Stapels und der ersten Karte des zweiten Stapels definiert wird, wurde wie selbstverständlich vorgeschlagen. Die Möglichkeit, jeden anderen Zwischenwert anstelle des arithmetischen Mittelwerts als Median zu nehmen, wurde von den Lerngruppen überhaupt nicht ins Auge gefasst.

Minimum, Maximum und Median können über relativ einfache Anweisungen über den Umgang mit Kartenstapeln bestimmt werden. „Kann ein Computer den Median berechnen? Gibt es hierfür eine For-mel?“ Diese Fragen führten zur Suche nach einer anderen Möglichkeit zur Bestimmung des Medians. Schlüsselfrage war, wie zu bekannter Anzahl n an Karteikarten die Nummer der Karte(n), auf der (denen) die für den Median wesentlichen Werte verzeichnet ist (sind), berechnet werden kann. Übungen mit Karteistapeln unterschiedlicher Größe erbrachten die Informationen, so dass wir für den Median der Zahlen x_i mit $1 \le i \le n$, hier $x_{0,5}$ genannt, folgende Definition formulieren konnten:

$$x_{0,5} = \begin{cases} \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} & \text{n gerade} \\ x_{\frac{n+1}{2}} & \text{n ungerade} \end{cases}$$

Die Veranschaulichung mit quer gestellten Karteikarten, die Definition und die Berechnung des Medians wurden ernsthaft auf die Probe gestellt, als die Frage nach dem 1. und 3. Quartil aufgeworfen wurde. Ich habe die einfachen Möglichkeiten gewählt und das 1. Quartil als Median der ersten Hälfte, das 3. Quartil als Median der zweiten Hälfte des Kartenstapels definiert. Damit die Lernenden alle Halbierungen mit ihren Übungskarteien durchführen konnten, musste jede Karteikarte in vierfacher Ausfertigung vorliegen, es sei denn, die Anzahl der Karteikarten ist bereits eine durch 4 teilbare Zahl. Als Vorschrift zur Berechnung des 1. Quartils der Zahlen x_i mit $1 \le i \le n$, hier $x_{0,25}$ genannt, und des 3. Quartils, hier $x_{0,75}$ genannt, haben wir folgende Formeln erarbeitet:

$$x_{0,25} = \begin{cases} \dfrac{x_{\frac{n}{4}} + x_{\frac{n}{4}+1}}{2} & \text{n durch 4 teilbar} \\ x_{\frac{n+4-r}{4}} & \text{n sonst} \end{cases}$$

Hierbei bezeichnet r den Rest, der bei der Division von n durch 4 entsteht.

$$x_{0,75} = \begin{cases} \dfrac{x_{3\cdot\frac{n}{4}} + x_{3\cdot\frac{n}{4}+1}}{2} & \text{n durch 4 teilbar} \\ x_{\frac{3n+r}{4}} & \text{n sonst} \end{cases}$$

Die ersten Zeilen beider Beschreibungen waren analog zur Beschreibung des Medians so zu vermuten und konnten anhand der Erfahrung mit kleinen Datensätzen und nach der Vorarbeit beim Median auch relativ einfach entdeckt und dann aufgeschrieben werden. Die zweiten Zeilen machten schon mehr Mühe, bis schließlich erkannt wurde, dass es genau der Viererrest war, um den der Zähler verringert beziehungsweise vergrößert werden musste.

Für diese zweite Zeile machten einige Schüler einen anderen Vorschlag, auf den ich nun eingehen möchte. In Anlehnung an die in deutschsprachigen Tabellenkalkulationen vorhandenen Funktionen definierten sie eine Anweisung abrunden(x). Sie formulierten in den zweiten Zeilen der oben angeführten Rechenvorschriften abrunden $(\frac{n}{2})+1$ beim Median, abrunden $(\frac{n}{4})+1$ beim 1. Quartil und beim 3. Quartil abrunden $(\frac{3n}{4})+1$. „Kann man das in der Mathematik auch machen?“, fragten diese Schüler, die sich in einer Tabellenkalkulation schon recht gut auskannten. Hier war nun der Ort die Gauss-

Klammer [x] zunächst nur für nicht-negative Zahlen einzuführen, einfach als [x]:= abrunden(x). Da die Formulierungen mit Hilfe der Gauss-Klammer kürzer waren als die Versuche mit abrunden(x), wurde [x] als Hilfe und Erleichterung akzeptiert.

Zum Schluss vereinigten die Lernenden die Karteikarten für die fünf Kennzahlen zu einer einzigen. Für mich war dies eine interessante und überzeugende Darstellung der „Verdichtung“ oder Reduktion der Information des gesamten Datensatzes auf die fünf Kennzahlen.

Nun festigten die Lernenden den Umgang mit diesen Grundbegriffen beim Umgang mit ihren Karteikarten, mit Stengel-Blätter-Diagrammen sowie bei der Erstellung von Boxplots. Jeder Lernende sollte die Daten von etwa zehn Automodellen besorgen. Die Arbeit wurde so aufgeteilt, dass die Daten von Autos aus elf Ländern zusammengetragen wurden. Jeder Schüler notierte von jedem Modell seines Aufgabenbereichs alle gefragten Daten auf einer eigenen Karteikarte. Schon bei der Datensammlung ergaben sich genug Anknüpfungspunkte für Gespräche unter den Lernenden, zum Beispiel darüber, wie fehlende Daten beschafft werden können oder darüber, was unter normalen Gebrauchs-Pkw zu verstehen ist. Zunächst erstellte jeder Lernende von seinen Modellen ein Stengel-Blätter-Diagramm, bestimmte die Kennzahlen und zeichnete einen Boxplot. Mit diesen optischen Hilfen konnten viele Vergleiche zwischen den Modellen einzelner Firmen angestellt werden, zum Beispiel zwischen VW und Opel oder Fiat. Mit solchen Vergleichen wurde die Unterrichtseinheit in den 8. Klassen beendet. Die Grundkurse in der gymnasialen Oberstufe führten alle Daten in einem einzigen Arbeitsblatt einer Tabellenkalkulation zusammen, so dass nun die oben dargestellten Boxplots zu den einzelnen Merkmalen erstellt werden konnten.

Am Beispiel der Autos aus Italien soll die Erstellung eines Boxplots der beurteilenden Statistik betrachtet werden. Es liegen 24 Daten zur Leistung in kW vor, die der Größe nach geordnet angegeben werden: 29, 40, 40, 44, 54, 59, 76, 76, 76, 83, 83, 88, 103, 108, 110, 114, 129, 132, 150, 209, 247, 280, 357, 390. Es gilt: Minimum = 29; 1. Quartil = 67,5; Median = 95,5; 3. Quartil = 141; Maximum = 390. Das Rechteck vom 1. Quartil zum 3. Quartil kann nun gezeichnet und darin die Lage des Medians markiert werden. Weiter gilt: R = 73,5. 1. Quartil - 1,5R ist negativ, also gibt es keine möglichen Ausreißer bei kleinen Leistungen. Nun können wir den unteren Whisker vom Minimum bis zum 1. Quartil zeichnen. 3. Quartil + 3R = 361,5, also wird die größte Leistung 390 kW mit „o“ gekennzeichnet. 3. Quartil + 1,5R = 251,25. Die zwei Leistungen (280 und 357 kW werden mit „*“ bezeichnet. Vom 3. Quartil bis zu 247 kW wird nun der obere Whisker gezeichnet, der Boxplot ist fertig. Wesentlich schneller geht dies unter Benutzung von Software, aber nach den obigen Überlegungen und Rechnungen können wir nachvollziehen, wie die erste Abbildung in Abschnitt 3 zustande gekommen ist. Beim zweiten Boxplot zur Leistung ohne Sportwagen müssen wir berücksichtigen, dass die drei größten Leistungen (280, 357, 390 kW) wegfallen.

Am Ende der Unterrichtseinheit entwickelten die 8. Klassen eine besondere Strategie: Die Bereiche der Whiskers vom 3. Quartil bis $x_{0,75}$ + R und vom 1. Quartil bis $x_{0,25}$ - R wurden farbig eingezeichnet. Im jeweils überstehenden Teil wurde nach möglichen Ausreißern gefahndet. Die Begründung der Lernenden war sehr pragmatisch. So seien sie vorher oft auf außergewöhnliche Autotypen gestoßen, die sich stark von normalen Gebrauchs-Pkw unterschieden. In den Grundkursen der gymnasialen Oberstufe gingen die Lernenden anders vor. In Erinnerung an ihre Erfahrungen mit σ-Umgebungen (vgl. hier zum Beispiel Wirths 1996) formulierten sie: Alles, was sich um mehr als 2σ vom arithmetischen Mittelwert μ unterscheidet, wird als möglicher Ausreißer betrachtet. Folgerichtig ließen sie sich in der Tabellenkalkulation das Maximum, das 3. Quartil, den Median, das 1. Quartil, das Minimum, den arithmetischen Mittelwert μ, die Standardabweichung σ sowie $\mu + 2\sigma$ und $\mu - 2\sigma$ ausrechnen. Einige mit der Tabellenkalkulation besonders Vertraute ließen sich für jedes Merkmal alle Modelle außerhalb der jeweiligen 2σ-Umgebung um μ im Rechenblatt besonders hervorheben. Ihr Vorgehen begründeten sie so: Bei binomialverteilten Größen liegen in etwa 95 % aller Daten im Bereich zwischen $\mu + 2\sigma$ und $\mu - 2\sigma$, außergewöhnliche Werte, also seltene oder auch Ausreißer, seien vor allem in den restlichen ungefähr 5 % zu erwarten. Für andere Verteilungen wollten sie diese Regel als Näherung benutzen. In jedem Einzelfall wollten sie eine Untersuchung auf Ausreißer führen. Dadurch könne die Gefahr von Fehlentscheidungen bei allzu schematischem Vorgehen vermieden werden.

Vielleicht war diese Strategie auch eine Reaktion auf Softwareprobleme, über die sie sich beklagten. Die Lernenden nutzten die Tabellenkalkulation am Computer und ihren eigenen graphikfähigen Taschenrechner als Tutoren. Sie entdeckten dabei, dass die beiden Quartile nicht immer korrekt errechnet werden. Dass nicht wie oben beschrieben der Mittelwert zwischen den Werten benachbarter Karten gebildet, sondern irgendein anderer Zwischenwert genommen wird, darüber wollten die Lernenden noch hinwegsehen. Aber dass Excel und Quattro Pro bei den Werten 1, 2, 3, 4, 5, 6 das 1. Quartil als 2,25 und das 3. Quartil als 4,75 ausgeben, so dass $\frac{1}{3}$ der Werte unterhalb des 1. Quartils liegen, das wollten sie nicht hinnehmen. Dass beim TI-83 aus den Werten 1, 2, 3, 4, 5 das 1. Quartil als 1,5 und das 3. Quartil als 4,5 angegeben wird, so dass 80 % der Werte über dem 1. Quartil liegen, sei zwar schon besser, aber auch noch nicht korrekt. Mehr Sorgfalt bei der Erstellung von Software könne man von den großen Firmen schon erwarten. Schließlich sei die Berechnung der Quartile, wie die oben angeführten Überlegungen zeigen, kein Problem.

Beim Hantieren mit Karteistapeln kann man eine wichtige Eigenschaft des Median erfahren. Entfernt man eine Karteikarte oberhalb und eine unterhalb des Medians, ändert sich der Median nicht. Das gilt unabhängig davon, welchen Abstand die Werte auf den Karteikarten vom Median haben. Man sagt, der Median sei robust gegenüber Ausreißern, eine Eigenschaft, die man dem arithmetischen Mittelwert nicht zuordnen kann. Dies wird auf die Probe gestellt, wenn wir die beiden Boxplots zur Leistung in Abschnitt 3 betrachten. Bei den Autos aus Großbritannien liegt der Median bei allen 19 Modellen bei 120 kW, entfernt man die Sportwagen, sinkt er auf 96 kW. Die Erklärung für diesen Effekt ist, dass wir 3 Werte oberhalb und nur einen unterhalb des Medians entfernen. Wir müssen zudem berücksichtigen, dass die Bereiche vom 1. Quartil bis zum Median und vom Median zum 3. Quartil, Bereiche, in denen jeweils 25 % aller Leistungen liegen, relativ groß sind, Änderungen sich hier also viel drastischer auswirken können als bei kleineren Bereichsgrößen.

5 Abschluss

Ich habe mich bemüht, bei den Lernenden Ideen zur Ausgestaltung des Unterrichts zu provozieren, Anregungen Ernst zu nehmen und möglichst auch zu berücksichtigen. Darüber habe ich in diesem Aufsatz berichtet. Möglich wären auch andere Kriterien zur Auswahl der Autos gewesen. Statt Sportwagen auszuschließen, könnte man einen Höchstpreis zum Beispiel l00 000 DM vorgeben und so eine Auswahl treffen. Aber fehlt dann in den Boxplots nicht das „Salz in der Suppe“, die mit „o“ oder „*“ gekennzeichneten besonders auffälligen Autotypen? Auch eine Berücksichtigung der Zulassungszahlen, allerdings sehr aufwändig zu beschaffen, wäre denkbar. Man könnte auch Pkw mit Dieselmotor oder nur Modelle mit Automatikgetriebe untersuchen. All dies wurde in meinen Lerngruppen diskutiert, aber nicht weiter verfolgt. Ich habe mich sehr darüber gefreut, dass in den 8. Klassen und auch in den Pflichtauflagenabdeckerkursen der gymnasialen Oberstufe Schülerinnen und Schüler durch diese Problemstellung motiviert werden konnten.

Für solch eine Unterrichtsreihe sollte man gut 3 Wochen Unterricht einplanen. Es hat sich als günstig erwiesen, dass ich nach Vorstellung des Vorhabens und einer ersten Diskussion den Lernenden über eine Woche Zeit zum Sammeln von Material gelassen und in dieser Zeit eine andere Unterrichtseinheit abgeschlossen habe. In den 8. Klassen musste ich zunächst mein Augenmerk stark auf das Erfassen der Daten und das Erstellen der Karteikarten richten, während in den Pflichtauflagengrundkursen in der gymnasialen Oberstufe die Eingabe in den Computer, insbesondere das Zusammenstellen der insgesamt 150 Daten in einer Excel-Datei von den Lernenden in ihren Freistunden selbständig organisiert wurde.

Die Frage „Sind deutsche Autos anders als ausländische?“ ist hier bewusst nur am Rande gestreift worden. Die Boxplots sind aussagekräftig, die unterschiedlichen Kennzahlen (1., 3. Quartil, Median), die Länge der Whiskers, der Interquartilsabstand, die mit „*“ oder „o“ besonders gekennzeichneten Modelle sowie die Unterschiede zwischen den Boxplots mit Sportwagen und denen ohne diese Autos, all das reizt zu Diskussionen und zum Vergleich. Daraus ergeben sich viele Aussagen, mit denen die Unterschiede zwischen deutschen und ausländischen Pkw beschrieben werden können. Ist das nicht eine Herausforderung für jeden Leser oder jede Leserin, dies einmal selbst zu versuchen, oder - noch besser - die eigene Lerngruppe zu solchen Aussagen zu animieren? Oder wie wäre es mit anderen Gegenständen als Autos?

Für Kolleginnen und Kollegen, die mit der explorativen Datenanalyse EDA noch nicht vertraut sind, eine kleine Auswahl an Literatur zum Einarbeiten:

Biehler, R.: Explorative Datenanalyse. IDM Materialien und Studien Bd. 24, Bielefeld 1982

Borovcnik, M. / Ossimitz, G.: Materialien zur Beschreibenden Statistik und zur Explorativen Datenanalyse. Teubner, Stuttgart 1987

Engel, A.: Stochastik, Klett, Stuttgart 1987

Henze, N.: Stochastik für Einsteiger. Vieweg, Braunschweig 1997

Kütting, H.: Beschreibende Statistik im Schulunterricht. BI: Mannheim 1994

Nordmeier, G.: „Erstfrühling“ und „Aprilwetter“ - Projekte in der explorativen Datenanalyse. Stochastik in der Schule 9 (1989) 3, S. 21 - 42

Reichel, H.-C.: Wahrscheinlichkeitsrechnung und Statistik. Hölder-Pichler-Tempsky: Wien 1987

Trauerstein, H.: Daten erheben, bearbeiten und auswerten - Erfahrungen mit einem Reaktionstest bei Schülern und Studenten. Stochastik in der Schule 15 (1995) 1, S. 13 - 23

Der Mathematikunterricht Heft 1 / 1982, Heft 6 / 1991, Heft 4 / 1997

Benutzte oder zitierte Literatur

Blankenburger, S.: Student Systat. Thomson Publ. Comp.1995

Wirths, H.: Schätz- und Prüfverfahren. Mathematik in der Schule Heft 11/ 1996, S. 596 - 607, 610 - 611

Anschrift des Verfassers:
Helmut Wirths
Starenweg 22
26131 Oldenburg
helmut.wirths@uni-oldenburg.de