استفاده از نمایش هندسی داده‌ها در تحلیل اطلاعات

تحقیق و نگارش: دکتر حمید ضرغام بروجنی

مقدمه

نگرش آمارشناسان به تحلیل داده‌ها، طی سی سال گذشته، دستخوش یک انقلاب شده است. این انقلاب را به درستی مرهون نشر اثر تاریخی جان توکی[1] تحت عنوان «تحلیل اکتشافی داده‌ها»[2] می‌دانند. هسته اولیه این انقلاب در مجموعه مقالات نوشته شده به وسیله توکی، همفکران و دانشجویان او شکل گرفت. این انقلاب با مقاله سال ۱۹۶۲ توکی تحت عنوان «آینده تحلیل دادها»[3] به جهت‌گیری روشنی دست یافت. محتوای این مقاله، به نیاز مشخص بسیاری از تحلیلگران داده‌ها که ابزار سنتی آمار ریاضی را برای تحلیل داده‌های خود کافی نمی‌دیدند پاسخ داد. پاسخ او برای آگاهی یافتن از جهات مختلف یک مسئله واقعی، بهره‌گیری از روشهای تحلیل اکتشافی داده‌ها بود.

امروزه تحلیل اکتشافی داده‌ها یک گام ضروری برای حل بسیاری از مسایل تحلیل

1- John Tukey
2- Exploratory Data Analysis
3- The Future of Data Analysis

آماری در علوم اجتماعی به طور اعم و مدیریت، به طور اخص بشمار می‌رود و فنون نمایش هندسی (تصویری) دادها نقش کلیدی این گام را بر عهده دارد. توسعه تحلیل اکتشافی داده‌ها با توجه به ضعف کاملاً آشکار روشهای آماری کلاسیک در تحلیل بسیاری از داده‌های مربوط به مسایل پیچیده اجتماعی، اجتناب‌ناپذیر بود. روشهای آماری کلاسیک مبتنی بر مفروضات خاصی هستند که ممکن است قابل حصول نباشند. به علاوه داده‌های ناجور[1] یا اندازه‌های پرت[2] که برای فهم فرایند تغییرات تحت مطالعه، اهمیت تعیین کننده دارند در این روشها منزلت ویژه‌ای ندارند. در نتیجه به فنونی نیاز بود که در آنها رعایت محدودیت های فرضی کمتری ضرورت داشته باشد.

با این وجود، تحلیل اکتشافی داده‌ها در پاره‌ای موارد به محاسباتی نیاز دارد که انجام آنها بدون در دسترس بودن رایانه عملی نیست. بنابراین انقلاب در ابزار محاسبه، یک جزء اساسی انقلاب در تحلیل داده‌ها بشمار می‌رود.

مرور تاریخی موضوع

حرکت تکاملی، از جدول به نمودار تحولی شگرفت در تحلیل داده‌ها است؛ مع الوصف آشنایی زیاد با نمودارها مانع تحسین شایسته از این دستاورد مهم شده است. بافت نگار، نمودار دایره‌ای، نمودار خطی و نمودار پراکنش، اساس آمار ترسیمی را تشکیل می‌دهند. رسم این نمودارها برای تحلیلگران، امکان درک اشکال مختلف توزیع داده‌های کمی را به صورتی اساساً متفاوت با جدول سازی آنها فراهم آورده‌اند. در واقع، بیان دوباره داده‌ها در قالب تصویر، یکی از تکامل یافته‌ترین توانمندی‌های پردازش اطلاعات در انسان، یعنی توانایی بازشناسی، طبقه‌بندی و به خاطر آوری الگوهای بصری را به یک سرمایه علمی تبدیل نموده است. «نمودارها به طور خاص از آن جهت مؤثرند که توانایی‌های طبیعی ادراکی، شناختی و به خاطر آوری انسان را مورد بهره‌برداری قرار می‌دهند» (Kosslyn, 1989).

1- Spotty data 2- Outliers

امروزه رسم نمودارها بسیار ساده و معمولی به نظر می‌رسند ولی باید اذعان کرد که ابداع آنها نه ساده بود و نه مشهود. ایده تبدیل انبوه داده‌ها به نمودار، نه به اذهان هخامنشیان یا یونانیان باستان خطور کرد و نه حتی به ذهن ریاضیدانان بزرگ قرن هفدهم. ابزار اصلی و مهم ابداع تقریباً همه نمودارهای آماری، هندسه دکارت (که در سال ۱۶۱۳ میلادی نشر یافت) است (Fox, 1990). هندسه تحلیلی جدید برای فیلسوفان و مهندسان ابزار پر جذبه‌ای برای شناسایی رفتار توابع ریاضی و گاهی مواقع، نمایش روابط بین متغیرهای کالبدی بشمار می‌رفت (Tilling, 1975).

ابداع بیشتر نمودارهای پر کاربرد آماری امروزین (از جمله نمودار دایره‌ای، نمودار میله‌ای و هیستوگرام) به ویلیام پلیفیر (William Playfair) مهندس اسکاتلندی نسبت داده شده است (Playfair, 1801). او که بعدها به اقتصاد روی آورد؛ در دوران کودکی تحت آموزش برادرش (استاد ریاضیات دانشگاه Edinburgh) با هندسه دکارت[1] آشنا شد. در جوانی، درگیر ترسیم نمودارهایی به منظور نمایش عملکرد موتورهای بخار، در یک شرکت بود. بنابراین اسباب فنی لازم برای ابداع نمودارهای آماری برای او مهیا بود. پس از آن که به لندن آمد و بیشتر درگیر دنیای تجارت و بازرگانی شد، شیوه‌های تصویری مختلفی برای نمایش داده‌های اقتصادی ابداع کرد.

از آن زمان تا هنگامی که نمودارها در گزارش‌های علمی مورد استفاده قرار گرفتند مدتی به طول انجامید. به طوری که تا قرن نوزدهم در هیچیک از گزارشات علمی استفاده شایسته‌ای از هیچ نوع نموداری نشده بود و پس از آن نیز پیشرفت، بسیار کند بود (Shields, 1937). این کندی، نه تنها از عدم آشنایی کاربران با این نمودارها ناشی می‌شد بلکه مشکلات فنی تولید آنها نیز مانعی برای احتراز از کاربرد آنها به شمار می‌رفت. از سوی دیگر، چاپ نمودار در یک کتاب یا مجله، تا پیش از توسعه و تکمیل فنون عکاسی، فرایندی مشکل و وقت گیر بود.

1- Cartesian Geometry

اهمیت اطلاعات تصویری در درک موضوع

انسان برای بازیابی و پردازش الگوهای دیداری دارای قدرت ویژه‌ای است که نظام چشم-مغز[1] خوانده می‌شود. برطبق مطالعات موجود، بخش اعظم قدرت پردازشگری مغز انسان به بررسی اطلاعات دیداری اختصاص یافته است (Chambers, 1983). بسیار نادرند کسانی که بخواهند با این واقعیت که بینائی، کیفیت غالب مرکز حسی انسان است مخالفت کنند. از یک مجموعه داده‌هائی که در قالب دیداری به نمایش گذارده می‌شوند اغلب جزئیاتی را می‌توان دریافت که در داده‌های به شکل جدول غیر قابل مشاهده‌اند.

یک مثال تاریخی فراموش ناشدنی از اهمیت نمودارها به عنوان یک وسیله اکتشاف علمی، نمودار Hertzprug-Russell (H.R) است که با ثبت در بسیاری از لغت نامه‌های زبان انگلیسی، به اندازه کافی به آن اهمیت داده شده است (Random House, 1987). در سال ۱۹۱۳ در یک فراخوان انجمن سلطنتی علوم فضایی، Russell برای نخستین بار نموداری که بزرگی مطلق (یا روشنایی) ستاره‌ها را به عنوان تابعی از طبقه طیفی (یا حرارتی) آنها ترسیم کرده بود ارائه کرد. یک نگاه گذرا به این شکل پیچیده برای القاء تئوری نوین تحول کیهانی کافی بود. به حدی که اگر ادعا شود این نمودار صورت‌بندی تئوری تحول کیهانی را میسر ساخت، چندان اغراق نشده است. زیرا همین داده‌ها در سال ۱۹۰۵ توسط یک منجم اتریشی در قالب یک جدول ارائه شده بود ولی اطلاعات آن برای دانشمندان به حد لزوم بر انگیزاننده نبود (Spence, 1990).

به طور خلاصه می‌توان گفت از کاربرد روشهای ترسیمی داده‌ها چهار هدف مکمل زیر قابل تحصیل است:

- نخست آنکه، این روشها برای اکتشاف محتوای یک مجموعه داده‌ها مفیدند. نمودارهای آماری برای پاسخ به سؤالاتی درباره توزیع متغیرها در یک تحلیل (مثلاً شکل توزیع، دامنه تغییرات، مقادیر پرت یا مشاهدات ناجور) بسیار سریع‌تر از روشهای سنتی تلخیص داده‌های آماری که اغلب مبتنی بر مفروضات خاص هستند پژوهشگر را

1- eye-brain system

هدایت می‌کنند. به عبارت دیگر نمودارهای آماری فهم ما را از اطلاعات موجود در درون انبوه داده‌های خام تسهیل می‌کنند.

● دوم، آنکه نمودارهای آماری برای پیداکردن ساختار در داده‌ها به‌کار می‌روند. تفاوت عمده آنها از روشهای سنتی در آنست که می‌کوشند چارچوبهای فرضی محدود کننده درباره طبیعت این ساختارها تدوین نکنند. مثلاً در یک تحلیل رگرسیون ساده، یک متغیر وابسته به عنوان یک ترکیب خطی از مجموعه متغیرهای مستقل در نظر گرفته می‌شود. که به خودی خود یک گزاره انعطاف‌ناپذیر از اشکال مختلف ارتباط این متغیرها با هم است. در حالی که رسم نمودارها نشان می‌دهد چه نوع الگوی مرتب کننده‌ای به رابطه تغییرات این داده‌ها با هم قابل برازش است. و سپس تابع ریاضی بیان کننده این رابطه ساخته می‌شود. بنا به گفته توکی «این گونه روشها برای آن که با یک نگاه به دادها در یابیم که چه می‌خواهند بگویند کاربرد دارند» (Tukey, 1977).

● سوم، آنکه از نمایش تصویری داده‌ها برای بازبینی مفروضات مدلهای آماری استفاده می‌شود. می‌دانیم که کاربرد اغلب روشهای تحلیل آماری داده‌ها مستلزم برقراری مفروضات خاصی در مورد فرایند تولید داده‌ها است. اگر این مفروضات تحقق نیابند صحت هر نتیجه‌ای که از این تحلیل به دست آید در معرض تردید است. اگر چه همه دستورالعمل‌های آماری خوب در برگیرنده آزمونهای تشخیصی برای بازبینی مفروضات خود هستند اما استفاده از این آزمونها بسیار اندک است. مشاهده یک تصویر از توزیع داده‌ها جنبه‌های مشکل زای برازش یک مدل را به سهولت مشخص می‌کند. بنابراین، نمودارها محقق را برای استفاده ثمربخش‌تری از ابزارهای مدل‌سازی آماری هدایت می‌کنند.

● بالاخره، آنکه نمودارهای آماری برای انتقال نتایج یک تحلیل، بسیار مفیدند. همچنان که اشاره شد نظام فرایند دیداری انسان وسیله شگفت آوری برای فهم اطلاعات پیچیده است. معهذا، نمایش تصویری داده‌ها و مدلهای آماری در نشریات علمی (به ویژه علوم انسانی) به حد شگفت آوری نادرند.

ویژگیهای نمایش تصویری داده‌ها

نمودارها دو کاربرد اساساً متفاوت دارند:

- برقراری ارتباط اطلاعاتی با مخاطبین؛
- تحلیل داده‌ها و کشف جزئیات آنها.

اولین کاربرد، همگانی است و بیشتر، نمودارهای ساده و خوش ظاهر را شامل می‌شود. این قبیل نمودارها آخرین مرحله فرایند تحلیل داده‌ها را نمایش می‌دهند؛ و لذا معمولاً عرضه کننده آمارهای مجملند تا داده‌های اصلی. به این جهت، نکاتی که معرفی می‌کنند معمولاً خیلی ناچیز است. این قبیل نمودارها برای کاربرانی که تخصص آماری ناچیز دارند تهیه می‌شود و از این روی، ساده سازی قالب ظاهری و محتوا، در آنها یک گرایش عمده است. زیرا، در صورتی که کسب اطلاع از این نمودارها مشکل باشد، خود به یک معضل بدل می‌شوند تا وسیله‌ای برای فهم موضوع.

باید اذعان کرد که رسم نمودارهای تحلیلی داده‌ها، یک موضوع تخصصی است. این مجموعه در برگیرنده نمودارهایی است که حقایق پنهان در ورای داده‌ها را به نمایش می‌گذارند. در حالی که نمودارهای مورد استفاده برای تبادل اطلاعاتی ساده‌اند، نمودارهای تحلیلی کاملاً تخصصی بوده و ممکن است پیچیده یا مبهم به نظر برسند. نمودارهای تحلیلی علاوه بر نمایش وضعیت موجود داده‌ها، ابزار برجسته‌ای برای کشف جنبه‌های مهم ولی نامعلوم داده‌ها هستند. «این گونه نمودارها ما را مجبور به توجه به آن چیزهایی می‌کنند که هرگز انتظار دیدن آنها را نداریم» (Tukey, 1977).

انتخاب نمودار مناسب

هیچ ابزار آماری دیگری به قدرت یک نمودار خوب انتخاب شده، قادر به توضیح تغییرات متغیرها نیست. نظام چشم – مغز انسان، پیچیده‌ترین و تکامل یافته‌ترین دستگاه پردازش اطلاعات است. از طریق نمایشهای تصویری داده‌ها می‌توان از این نظام به نحو شایسته استفاده کرده نسبت به ساختار داده‌ها اطلاعات ارزشمندی به دست آورد. یکی

از مهمترین ویژگیهای نمودارها آنست که قادرند حجم زیادی اطلاعات کمّی را به سرعت منتقل نمایند. نظام چشم - مغز انسان می‌تواند به سرعت، اطلاعات بیکرانی را خلاصه کند و خصوصیات برجسته آنها را استخراج نماید؛ در عین حال با بهره‌گیری از یک نمودار مناسب، قادر است بر جزئیات آنها نیز متمرکز شود.

ارائه دستور العملهای مشخص برای دستیابی به نمودار تحلیلی مناسب، اگر نگوئیم غیر ممکن، ولی بسیار مشکل است زیرا به حد زیادی به ذوق و دانش تحلیلگر بستگی دارد. یک راهبرد در این زمینه، آنست که «وضعیتهای مهم مورد نظر در داده‌ها را با توجه به ویژگیهای برجسته کالبدی آنها که به سرعت قابل ملاحظه‌اند نشان دهید» (Cleveland, 1985).

توانایی تفسیر نمودارها نیز تا حد زیادی به میزان آموزش و تجربه کاربران آنها بستگی دارد. بیشتر افراد توانایی کافی برای درک نمودارهای خطی، نمودارهای میله‌ای، بافت نگار و مشابه آنها را دارند. در حالی که انتظار می‌رود تنها کسانی که دارای تحصیلات دانشگاهی هستند با نمودارهایی از قبیل: شاخه و برگ (Stem-and-Leaf diagrams)، جعبه و شاخک (Wisker-and-Box plots)، ریشه نگار (Quantile-Quantile plots) آشنا باشند. معهذا اگر چه، هر یک از این نمودارها دارای جایگاه ویژه‌ای در تحلیل داده‌ها هستند در عرضه آنها برای استفاده کنندگان کم اطلاع‌تر باید ملاحظات لازم را به کار گرفت.

برخی ملاحظات در ارائه نمودارها

نمودارهای یک متغیره، معمولاً برای بررسی توزیع متغیرها کاربرد دارند. نمودارهایی از قبیل بافت نگار و چند بر فراوانی از جمله اشکال پر کاربردتر این گونه نمودارها هستند. اما در سالهای اخیر نمایش شاخه و برگ، جعبه و شاخک و ریشه نگارها جای آنها را گرفته‌اند. مع الوصف برای انس گرفتن با این نمودارها و تحلیل دقیق داده‌ها به کمک آنها، به زمان بیشتری نیاز است. زیرا ظاهر برخی از این نمودارها ممکن است با اطلاعی که منتقل می‌کند کمی مغایر به نظر برسد. مثلاً هر یک از قسمتهای

«جعبه» در یک نمودار «جعبه و شاخک» همیشه، شامل یک چهارم مشاهدات است. این موضوع گاهی ممکن است موجب گمراهی تحلیلگر شود. به ویژه وقتی توزیع، نامتقارن باشد، مستطیل کوچکتر نشان دهنده فشردگی بیشتر داده‌ها است. اما چون راه مستقیمی برای نشان دادن فشردگی، وجود ندارد ممکن است بیننده، تصور کند سطح کوچکتر، مربوط به مشاهدات کمتر است و در مورد جهت چولگی دچار اشکال شود. شاید اضافه کردن نوعی سایه - روشنی به نمودار بتواند این نقیصه را جبران کند.

ریشه نگارها نیز بافت نگارهایی هستند که با استفاده از ریشه دوم داده‌های تبدیل یافته رسم می‌شوند و به کمک آنها چولگی و کشیدگی توزیع، مورد بررسی قرار می‌گیرد.

تلخیص داده‌ها با روش «شاخ و برگ»

اطلاعات مهمی که از نمودار (جدول) شاخه و برگ داده‌ها می‌توان به سرعت به دست آورد عبارتند از:

الف) گروه بندیهایی که ممکن است در داده‌ها وجود داشته باشند آشکار می‌شود و از این طریق دسته بندیهای درونی تغییرات متغیرها را می‌توان آشکار کرد.

ب) تقارن و یا عدم تقارن توزیع داده‌ها و اندازه‌هایی که خیلی بزرگ و یا خیلی کوچک باشند مشخص می‌شود.

ج) مقادیر فراوانی‌های خیلی زیاد و یا خیلی کم مشخص می‌شود.

برای آشنایی با نحوه ساخت این نمودارها می‌توان به کتابهای آماری مراجعه نمود اما برای آنکه استفاده‌های ممکن از روشهای نوین تلخیص تصویری داده‌ها عملاً نشان داده شود به مثالهای زیر توجه کنید.

جدول ۱- از روش شاخه و برگ برای نمایش اطلاعات موجود در مورد مدل و نام اتومبیل های عرضه شده در آگهی‌های فروش اتومبیل در تاریخ ۱۰ و ۱۱ اسفند ۱۳۶۰ روزنامه اطلاعات استفاده شده است (آگهی‌های تکراری ۱۱ اسفند را حذف کرده‌ایم). جدول ۱- ب نتیجه را وقتی که اتومبیلها را بر حسب حروف الفبا و سال ساخت آنها

مرتب کرده‌ایم نشان می‌دهد. توجه کنید که تا حد ممکن از حروف مناسب برای نمایش هر اتومبیل استفاده شده است. اما هر جا که استفاده از حروف، کد را خیلی پیچیده کرده از کد یک رقمی استفاده گردیده است. این کدها را در جدول ۱- الف مشاهده می‌کنید.

روشن است که اتومبیلهای پیکان و رنوی مدل پایین‌تر در بازار، زیاد عرضه نمی‌شود. همچنین اکثر اتومبیلهای خارجی که برای فروش عرضه می‌شوند مدلهای قدیمی می‌باشند. می‌توان حدس زد که اتومبیلهای تریومف، فولکس واگن و یا اتومبیلهای خارجی که در شاخه «بدون سال» مشاهده می‌شوند، مدل قدیمی باشند.

اکنون یک دسته از داده‌ها را در نظر بگیرید که هر عنصر آن به صورت یک جفت مرتب عدد (X,Y) باشند که در آن X و Y به ترتیب مقادیر متغیر یک جامعه را نمایش می‌دهند. در مثال زیر مشاهده می‌کنیم که برای خلاصه و تفسیر کردن چنین داده‌هایی چگونه می‌توان از نمایش شاخه و برگ استفاده کرد.

جدول ۲- الف مساحت و جمعیت ۲۴ استان کشور را (بر اساس سالنامه آماری سال ۱۳۶۵) نشان می‌دهد. نمایش شاخه و برگ این داده‌ها در جدول ۲- ب مشاهده می‌شود. باید توجه کنید که جفت (۵۳۳و۹) در شاخه‌ای که با «۱» شروع می‌شود نماینده جفت (۵۳۳و۱۹) مربوط به استان تهران است. به عبارت دیگر از روش شاخه و برگ ساده برای اولین عضو هر جفت استفاده گردیده و دومین عضو، به طور کامل نمایش داده شده است. دقت کنید که نمایش فوق، چگونه به استانهایی با تراکم جمعیت خیلی زیاد و یا خیلی کم اشاره می‌کند.

تلخیص‌های آسان - عددی و نموداری

۱- تلخیص چند عددی

اگر بخواهیم خصوصیات مهمی از داده‌ها که معرف شکل کلی تغییرات آنها است، به دست آوریم، یک راه مناسب آن است که این خصوصیات را در قالب چند عدد نمایش دهیم. اعدادی که به راحتی نیز قابل درک باشند. روشن است که اعداد به دست آمده

باید مجموعاً اطلاعات مفیدی ارائه دهند.

لازم به تذکر است که از تلخیص چند عددی نباید توقع زیادی داشت. زیرا نمی‌توان انتظار داشت که یک روش ساده و استاندارد که بتواند تمام جزئیات را ارائه دهد وجود داشته باشد. حتی گروه بندیهایی که در روش شاخه و برگ مشخص می‌شوند ممکن است در چنین تلخیص ساده‌ای منعکس نگردد. لذا برای اطلاع از جزئیات توزیع، مراجعه به روش شاخه و برگ و یا سایر روشهای مشابه ضروری باشد.

اگر قرار باشد چند عدد که به راحتی قابل محاسبه بوده و اطلاعاتی در مورد روند کلی داده‌ها بدهند بیابیم یقیناً بیشترین مقدار متغیر و کمترین مقدار آن و میانه را باید در نظر گرفت. اما این سه عدد، به تنهایی اطلاعات کافی در اختیار ما نمی‌گذارند، بنابراین لازم است دو عدد دیگر که آنها را لولا (چارک اول و چارک سوم) می‌نامند، معلوم کرده، یک «تلخیص پنج عددی» بسازیم. طبیعی است که این لولاها را باید بین نقاط حدی تغییرات (کوچکترین اندازه و بزرگترین اندازه) و میانه، در نظر بگیریم. قبل از محاسبه میانه و لولاها باید با مفاهیم «رتبه» و «عمق» هر مشاهده در داده‌ها آشنا شویم.

جدول ۳- الف مساحت ۱۷ دریاچه بزرگ دنیا را نشان می‌دهد. در جدول ۳- ب این دریاچه‌ها بر حسب مساحت، از کوچکترین به بزرگترین، مرتب شده‌اند. ستون «رتبه ↑» رتبه هر مشاهده را هنگامی که رتبه کوچکترین مشاهده برابر «۱» و رتبه بزرگتر برابر «۱۷» است نمایش می‌دهد. ستون «رتبه ↓» نیز رتبه هر مشاهده را نشان می‌دهد. ولی این بار بزرگترین مشاهده با رتبه «۱» و کوچکترین آنها با رتبه «۱۷» مشخص شده‌اند. کوچکترین رتبه‌ای که به هر مشاهده می‌توان نسبت داد «عمق» آن مشاهده نامیده می‌شود. توجه کنید که عمق هر مشاهده در ستون «عمق» مشخص گردیده است. عمق نقاط مربوط به کوچکترین و بزرگترین مقادیر متغیر برابر «۱» و عمق میانه برابر نصف مجموعه رتبه حداقل و رتبه حداکثر یعنی برابر ۹ = (۱۷+۱) ۰/۵ است. در حقیقت همانگونه که در مثالهای بعد مشاهده خواهید کرد عمق میانه همیشه برابر خواهد بود با:

۰/۵ (۱ + کل فراوانی)

معمولاً این پنج عدد را به صورت ساده‌ای که در جدول ۳- ث مشاهده می‌کنید نمایش می‌دهیم کل فراوانی یعنی «۱۷ #» در گوشه راست بالا و عمق نقاط مربوط به کوچکترین و بزرگترین مقادیر متغیر، میانه و لولاها در سمت راست به ترتیب با «ف- ۱»، «م - ۹» و «ل - ۵» نمایش داده شده‌اند. در نمایش عددی داده‌ها رعایت نکات زیر ضروری است:

الف: اگر عمق میانه یک عدد صحیح باشد، عمق لولا برابر است با:

(عمق میانه + ۱) ۰/۵

ب: اگر عمق میانه یک عدد کسری باشد، ابتدا مقدار «۰/۵» آن را حذف کرده و سپس قسمت الف عمق لولا را محاسبه می‌کنیم.

به اشکال ۳- پ و ۳- ت دقت کنید. این نمودارها نمایشهای واسطه‌ای برای نمودار عددی ۳- ث به شمار می‌روند. در نمایش ۳- پ ردیف داده‌ها از کمترین مقدار تا میانه و از میانه تا بیشترین مقدار شکسته شده است و در نمایش ۳- ت ردیف همان داده‌ها در سه نقطه چارک اول، میانه و چارک سوم شکسته شده است. هر یک از این نمایش‌ها اطلاعات قابل ملاحظه‌ای از تغییرات متغیر را آشکار می‌سازد.

۲- نمودار نقطه‌ای

معمولاً از تلخیص ۵- عددی و نمودار جعبه و شاخک مربوط به آن، روند کلی داده‌ها را می‌توان دریافت. ولی نباید انتظار داشته باشیم که چنین نمودار ساده‌ای بتواند جزئیات داده‌ها را نیز آشکار کند.

جدول ۴- الف ارتفاع بلندترین قله‌های ۴۰ کشور آسیا را نشان می‌دهد. همان گونه که دیده می‌شود ۷ کشور قله‌های مرتفعی ندارند. نمایش شاخه و برگ این داده‌ها را در جدول ۴- ب مشاهده می‌کنید. سه گروه را در این داده‌ها می‌توان تشخیص داد و انتظار داشت که این داده‌ها، دارای مقادیر دور و پرت باشند. ولی همانگونه که نمودار جعبه و شاخک استاندارد (جدول ۴- پ) نشان می‌دهد حتی قله اورست با ارتفاعی برابر

۸۸۴۸ متر نیز یک مقدار «دور» محسوب نمی‌شود.

روشن است که باید این داده‌ها را به طریق دیگری بررسی کرد. یکی از این روشها نمایش داده‌ها توسط «نمودار نقطه‌ای» است (نمودار ۴- ت). در چنین نموداری برای هر مقدار از داده‌ها نقطه‌ای روی یک خط عمودی رسم می‌کنیم. اگر داده‌ها در یک نقطه انباشته شده باشند نقاط را در طرف راست و یا چپ، به شکل افقی قرار خواهیم داد. همانگونه که ملاحظه می‌کنید نمودار نقطه‌ای، وجود سه گروه را به وضوح نشان‌می‌دهد.

با کمی دقت مشاهده می‌شود کلیه قله‌هایی که ارتفاعی بیش از ۷۰۰۰ متر دارند در سلسله جبال هیمالیا قرار دارند. اگر این قله‌ها را از سایر قله‌ها جدا کنیم می‌توان داده‌ها را مطابق جدول ۴- ث نمایش داد. توجه کنید که این نمایش شامل دو نمودار جعبه و شاخک می‌باشد که یکی برای قله‌های سلسله جبال هیمالیا و دیگری برای سایر قله‌ها است. مشاهده می‌کنید که این بار دوقله در دومین گروه به عنوان مقادیر «دور» مشخص می‌شوند.

البته ممکن بود برای این داده‌ها، سه نمودار جعبه و شاخک رسم کنیم (یک نمودار برای هر گروه) ولی از آنجایی که قله‌هایی که ارتفاع آنها بین ۵۰۰۰ و ۶۰۰۰ متر است هیچ‌وجه مشترکی به غیر از ارتفاع نزدیکشان به یکدیگر ندارند رسم سه نمودار، چندان منطقی به‌نظر نمی‌رسد.

کلام آخر

هدف از آنچه در این مقاله ارائه شد آشنایی خوانندگان ارجمند با وجه ارزشمند تحلیل تصویری داده‌ها بود. اهمیت تحلیلهای تصویری، اغلب به جهت تمایل مسلط تحلیلگران داده‌های کمی به کاربرد فرمول‌های پیچیده ریاضی، نادیده گرفته می‌شود. در حالی که استفاده از نمودارها در بسیاری موارد پاسخی به مراتب کاملتر از کاربرد روشهای کلاسیک بدست می‌دهد. نمودارهای معرفی شده در این مقاله، قسمت کوچکی از انبوه روش‌های تصویری است. استفاده مطلوبتر از این روشها مستلزم استفاده از رایانه و نرم‌افزارهای تخصصی در این زمینه است.

الف جدول (۱) تعداد اتومبیلهای فروشی آگهی شده در تهران

نام ماشین	کد
پیکان	پ
پیکان دو لوکس	پد
پیکان کار	پکا
پیکان جوانان	پج
وانت پیکان	۱
وانت نیسان	۲
وانت تویوتا لانکریزر	۳
مینی بوس	م
مینی بوس بنز	مب
آریا	آ
شورلت ایران رویال	ش
شورلت کاستم ایران	شکا
پژو ۴۰۴	پژ ۴
پژو ۵۰۴	پژ ۵
فولکس	ف
فولکس استیشن	فا
فیات ۱۲۴	فی
تویوتا کرونا	ت
کادیلاک	ک
فوردموستنگ	فو
داتسون ۱۲۰	د
تریمف	تر
بیوک گال	ب
لندرور انگلیسی	ل
ب.ام.و ۵۱۸	بم
رنو	ر
اتوموبیل خارجی	خ

	ب
۴۵	ل
۴۶	
۴۷	
۴۸	
۴۹	م آ
۵۰	ف پ
۵۱	
۵۱	فی
۵۳	مب
۵۴	۱ ۲ شکا ش ت پژ۴ ب
۵۵	فو شکا د پد پج بم
۵۶	م د ت پژ۵ پج پ بم
۵۷	ک پژ۵
۵۸	پد
۵۹	پکا پد پد پ
۶۰	پکا پکا پد پد پد پد
بدون سال	۳ فا رخ تر پکا پکا پد پ

جدول (۲) مساحت و جمعیت استانهای کشور

الف

نام استان	مساحت کیلومتر مربع	کل جمعیت
تهران	۱۹۱۲۵/۴	۵۳۳۱۱۶۶
مرکزی	۴۰۱۹۹/۰	۱۰۹۰۳۷۴
گیلان	۱۴۷۰۹/۰	۱۵۸۱۸۷۲
مازندران	۴۷۳۶۵/۰	۲۳۸۷۱۷۱
آذربایجان شرقی	۶۷۱۰۲/۴	۳۱۹۷۶۸۵
آذربایجان غربی	۳۸۸۵۰/۰	۱۴۰۷۶۰۴
کرمانشاه	۲۳۶۶۶/۵	۱۰۳۰۷۱۴
خوزستان	۶۴۷۰۱/۰	۲۱۸۷۱۱۸
فارس	۱۳۳۲۴۵/۰	۲۰۳۵۵۸۲
کرمان	۱۹۳۱۳۱/۰	۱۰۹۱۱۴۸
خراسان	۳۱۳۳۳۷/۲	۳۲۶۴۳۹۸
اصفهان	۱۰۴۷۰۰/۰	۲۱۷۶۶۹۴
سیستان و بلوچستان	۱۸۱۵۷۸/۰	۶۶۴۲۹۲
کردستان	۲۴۹۹۸/۰	۷۸۲۴۴۰
همدان	۲۰۱۷۲/۰	۱۰۸۸۰۲۴
چهارمحال و بختیاری	۱۴۸۲۰/۳	۳۹۴۳۵۷
لرستان	۳۱۳۸۳/۲	۹۳۳۹۳۹
ایلام	۱۹۰۴۴/۰	۲۴۶۰۲۴
بویر احمد و کهکیلویه	۱۴۲۶۱/۰	۲۴۴۳۷۰
بوشهر	۲۷۶۵۳/۰	۳۴۷۸۶۳
زنجان	۳۶۳۹۸/۳	۱۱۱۷۱۵۷
سمنان	۹۰۰۳۹/۰	۲۸۹۴۶۳
یزد	۵۶۸۹۶/۰	۳۵۶۸۴۹
هرمزگان	۶۶۸۷۰/۴۵	۴۶۲۴۴۰

	ب
۰ *	
۱	(۲۴و۴) (۲۴و۹) (۲۹و۴) (۱۵۸و۴) (۵۳۳و۹)
۲	(۱۱۱و۶) (۳۴و۷) (۱۰۸و۰) (۷۸و۴) (۱۰۳و۳)
۳	(۹۳و۱) (۱۴۰و۸)
۴	(۲۳۸و۷) (۱۰۹و۰)
۵	(۳۵و۶)
۶	(۳۱۹و۷)
۷	
۸	
۹	(۲۸و۰)
۱ **	(۶۶و۸۱) (۲۱۷و۰۴) (۱۰۹و۹۳) (۲۰۳و۳۳)
۲	
۳	(۳۲۶و۱۳)

جدول (۳) مساحت بزرگترین دریاچه‌های دنیا

الف

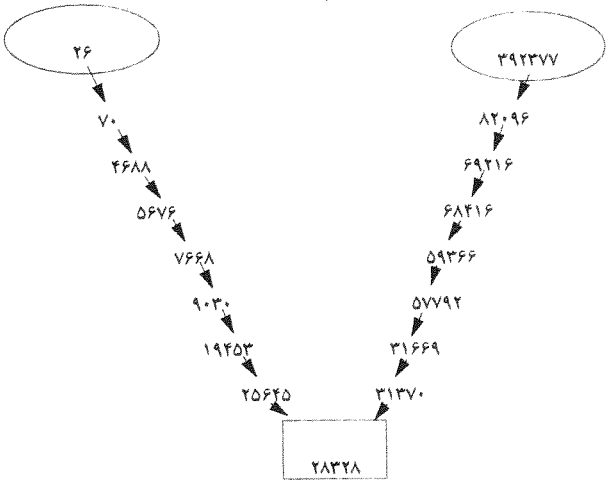
نام قاره	نام دریاچه	مساحت (کیلومترمربع)
آسیا....	خزر	۳۹۲۳۷۷
	آرال	۶۸۴۱۶
	بایکال	۳۱۳۷۰
آفریقا....	ویکتوریا	۶۹۲۱۶
آمریکای جنوبی	تی تی کاکا	۹۰۳۰
	نیکاراگوا	۷۶۶۸
آمریکای شمالی	سوپریور	۸۲۰۹۶
	هورن	۵۹۳۶۶
	میشیگان	۵۷۷۹۲
	خرس بزرگ	۳۱۶۶۹
	اسلیو بزرگ	۲۸۳۲۸
	اریه	۲۵۶۴۵
	انتاریو	۱۹۴۵۳
	مانیتوبا	۴۶۸۸
اروپا....	لُخ لوموند	۷۰
	ویندرمیر	۲۶
استرالیا	تورنس	۵۶۷۶

ب

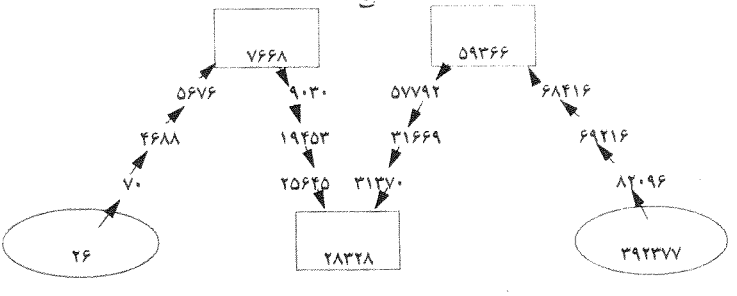
عمق	رتبه	رتبه	مساحت (کیلومترمربع)
۱	۱۷	۱	۲۶
۲	۱۶	۲	۷۰
۳	۱۵	۳	۴۶۸۸
۴	۱۴	۴	۵۶۷۶
۵	۱۳	۵	۷۶۶۸
۶	۱۲	۶	۹۰۳۰
۷	۱۱	۷	۱۹۴۵۳
۸	۱۰	۸	۲۵۶۴۵
۹	۹	۹	۲۸۳۲۸
۸	۸	۱۰	۳۱۳۷۰
۷	۷	۱۱	۳۱۶۶۹
۶	۶	۱۲	۵۷۷۹۲
۵	۵	۱۳	۵۹۳۶۶
۴	۴	۱۴	۶۸۴۱۶
۳	۳	۱۵	۶۹۲۱۶
۲	۲	۱۶	۸۲۰۹۶
۱	۱	۱۷	۳۹۲۳۷۷

جدول (۳)... دنباله

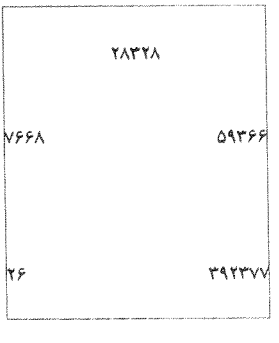
پ



ت



ث



(میانه) م – ۹

(لولا) ل – ۵

(نقاط حدی) ف – ۱

جدول (۴) مساحت بزرگترین دریاچه‌های دنیا

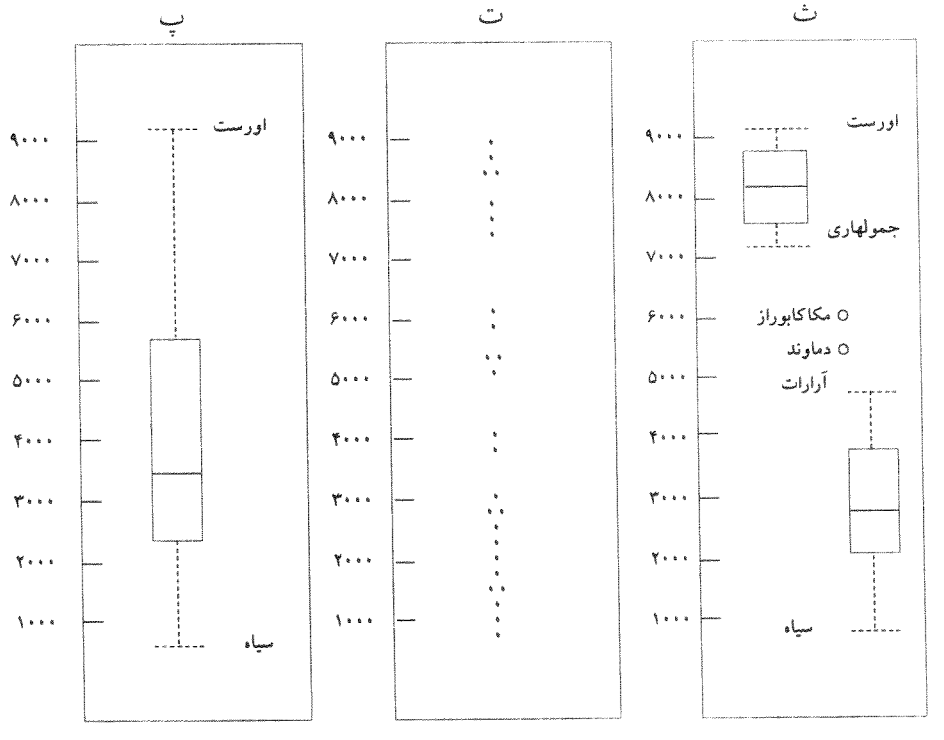
الف

نام کشور	نام قله	ارتفاع
اردن هاشمی	رام	۱۷۵۴
اسرائیل	خاروف	۱۰۱۶
افغانستان	شاه فولادی	۵۱۲۹
اندونزی	سوکارنو	۵۰۲۹
ایران	دماوند	۵۵۸۶
بحرین	ندارد	—
برمه	هکاکابو رازی	۵۸۸۱
بوتان	جمولهاری	۷۲۹۶
بنگلادش	ندارد	—
پاکستان	تنگاپربت	۸۱۲۵
تایلند	میانگ	۲۳۱۶
ترکیه	آرارات	۵۱۵۵
چین (تبت)	گدوین استن	۸۶۱۱
ژاپن	فوجی یاما	۳۷۷۵
سنگاپور	نی سون	۱۱۰۰
سوریه	طلعت موسی	۲۶۵۶
سری لانکا	پیدوروتالاگالا	۲۵۱۷
عراق	حاجی ابراهیم	۳۶۰۰
عربستان سعودی	رأس الصبح	۳۰۰۰
فیلیپین	ماتوتوم	۲۳۳۰

نام کشور	نام قله	ارتفاع
قبرس	ترودوس	۱۹۸۰
قطر	ندارد	—
کره جنوبی	شبری سان	۱۹۵۰
کره شمالی	موان مولونگ	۲۵۸۰
کامبوج	خولونگ یانی	۱۱۴۵
کویت	ندارد	—
لائوس	پوبیا	۲۸۲۰
لبنان	قرنتس سعودا	۳۰۸۶
مالزی	گونونگ بنوم	۲۱۴۰
مالدیو	ندارد	—
مغولستان	ندارد	—
نپال	اورست (جملونگما)	۸۸۴۸
ویتنام	فان سی پان	۳۱۴۳
هند	نند ادووی	۸۱۰۰
یمن جنوبی	اداران	۲۵۰۰
یمن شمالی	منار	۳۳۰۰
شوروی (آسیایی)	کمونیزم (تاجیکستان)	۷۴۹۵
مسقط و عمان	اختر	۳۰۰۹
شیخ‌نشینان خلیج فارس	ندارد	—
موریس	سیاه	۸۲۷

ب

۰***	۸۲۷
۱	۰۱۶، ۱۰۰، ۱۴۵، ۷۵۴، ۹۵۰، ۹۸۰
۲	۱۴۰، ۳۱۶، ۳۳۰، ۵۰۰، ۵۱۷، ۵۸۰، ۶۵۶، ۸۲۰
۳	۰۰۰، ۰۰۹، ۰۸۶، ۱۴۳، ۳۰۰، ۶۰۰، ۷۷۵
۴***	
۵	۰۲۹، ۱۲۹، ۱۵۵، ۵۸۶، ۸۸۱
۶	
۷	۲۹۶، ۴۹۵
۸***	۱۰۰، ۱۲۵، ۶۱۱، ۸۴۸



منابع و مآخذ

1- Fox, J. and J. Scott Long. (1990). "Modern Methods of Data Analysis". USA: SAGE Publications.

2- Jacoby, G. (1997). "Statistical Graphics for Univariate and Bivariate Data". USA: SAGE Publications.

3- Kosslyn, S. M. (1989). "Understanding Charts and Graphs. Applied Cognitive Psychology", 3, 185.

4- Lambert, J. H. (1779). "Pyrometrie". Berlin: Bebey Houde.

5- Plyfair, W. (1801). "Statitical Breviary". London: Wallis.

6- Shields. M. C. (1937). "The Early History of Graphs". American Physics Teacher, 5, 68.

7- Spence. I. and Lewandowksky, S. (1990). "Displaying proportions". Applied Cognitive Psychology, 3.

8- Tilling. L. (1975). "Early experimental graphs". The British Journal for the History of Science, 8, 193.

9- Tukey, J. W. (1972). "Some Graphics and Semigraphics Display". Iowa State University.

10- Tukey, J. W. (1977). "Exploratory Data Analysis". MA: Addison-Wesley.
۱۱- «سالنامه آماری کشور». (۱۳۶۵). مرکز آمار ایران.
۱۲- «سالنامه آماری کشور». (۱۳۷۵). مرکز آمار ایران.
۱۳- گروه کار پژوهش. (۱۳۶۳). «طرح آموزش آمار بدون پیش نیاز ریاضی». مرکز آمار ایران.