

От редактора: Говорим на языке диаграмм, или Как верно визуализировать данные в научной статье

Татьяна А. Баклашова¹, Айдар М. Калимуллин²

¹ Казанский (Приволжский) федеральный университет, Казань, Россия

E-mail: ptatyana2011@mail.ru

ORCID: <https://orcid.org/0000-0002-3038-6990>

² Казанский (Приволжский) федеральный университет, Казань, Россия;

Самаркандский государственный университет им. Шарофа Рашидова, Самарканд,
Узбекистан

E-mail: kalimullin@yandex.ru

ORCID: <https://orcid.org/0000-0001-7788-7728>

DOI: 10.26907/esd.20.1.01R

EDN: FHVSRK

Сегодня мы живем в мире данных, они окружают нас повсюду. Простые и сложные, разрозненные и четко организованные на основе ряда факторов, они помогают нам проверять гипотезы, выявлять закономерности, делать открытия (Larose & Larose, 2014). Будучи тесно связанными между собой, исследования и данные сопряжены с важнейшими этапами научного изыскания – сбором информации и ее анализом. Анализ данных направлен на выявление закономерностей и достижение ясности в исследуемом явлении.

В каждой науке данные играют свою роль. В образовании данные помогают улучшить качество обучения и учебного процесса (Bienkowski et al., 2012). Становится более доступной персонализация обучения: качественные и количественные показатели успеваемости и предпочтений ученика позволяют педагогам отбирать содержание, определять темп и методы обучения с учетом его особенностей. Кроме того, показатели позволяют прогнозировать успеваемость обучающихся, их отсев, если речь идет о вузе. Благодаря данным можно совершенствовать содержательный аспект: обновлять программу обучения, актуализировать фонд оценочных средств для обеспечения качества образования. Значимы данные и в вопросе управления образовательной организацией: по ним руководство способно следить за финансовым состоянием организации, планировать бюджет, осуществлять эффективную координацию всей системы. Данные связаны также с инновациями в образовании, позволяя разрабатывать новые онлайн-курсы, электронные учебники, приложения, направленные на обеспечение доступности и эффективности обучения.

Работа с данными предполагает их визуализацию. Существует большое количество рекомендаций, статей по вопросу графического представления данных (Maaten & Hinton, 2008; Glazer, 2011; Yau, 2024). Мы расскажем здесь об основных, наиболее значимых аспектах.

Итак, в научных работах правильная визуализация данных помогает лучше понять сложную информацию, касающуюся проведенного исследования. Одним из способов визуального представления и структурирования данных являются графики. Каковы их преимущества? Во-первых, они позволяют упростить (а не усложнить!) излагаемый материал, представить его в виде понятных читателю изображений. Во-вторых, при построении графика учитываются типы данных – в про-

тивном случае график не будет отражать реального процесса и запутает читателя. В-третьих, графики не должны отражать причины изменения данных, а если автор имеет целью представить именно их, то ему нужно воспользоваться иным методом визуализации. В-четвертых, в представлении результатов исследовательского труда удобнее работать с графиками, чем с фактической информацией, поскольку они в сжатом виде демонстрируют объемный массив данных, способствуя пониманию читателем тенденций по показателям.

Какова роль графического метода в иллюстрации взаимосвязей между данными в исследованиях? Здесь хотелось бы обратить внимание на ряд моментов. Во-первых, график помогает исследователю доказать свою точку зрения. Последняя однозначно выглядит убедительнее, когда рядом с табличными результатами имеется еще и графическое их представление. В этом случае читатель быстрее поймет вывод, который делает автор работы, не запутается в декодировании его расчетов. Во-вторых, график позволяет проанализировать большой массив данных, занимая при этом минимум места в статье. В мире издательского дела часто звучит идея о том, что передача большего количества деталей в наименьшем объеме слов и пространства – это искусство, которое можно и нужно практиковать с помощью графиков. В-третьих, графики помогают автору представить сложные данные в упрощенном виде, повышая уровень корректного восприятия мыслей автора читателем. В-четвертых, графики позволяют отразить сравнение данных. В том случае, если исследователь сравнивает в статье два и или более набора данных, которые представляют собой большое количество критериев и показателей, целесообразнее использовать, например, столбчатую диаграмму, чтобы читатель мог понять разницу в состояниях этих наборов данных.

Важно помнить, что графики не рекомендуется использовать при описании больших объемов данных – в этом случае целесообразнее прибегнуть к их краткому текстовому изложению. Также не нужно повторять данные, которые приводятся в графике, в текстовом формате – иначе нивелируется цель использования графика, которая была изложена нами выше, а именно передача большого количества данных в сжатом формате. Не следует использовать график, если данные не показывают какой-либо существенной тенденции, позволяющей автору доказать его гипотезу, ответить на исследовательский вопрос.

В большинстве случаев график строится таким образом, чтобы горизонтальная ось (ось ОХ) представляла независимую переменную, а вертикальная ось (ось ОУ) – зависимую переменную. Следовательно, если речь идет о временных промежутках, то они всегда откладываются по оси ОХ. График должен содержать как минимум заголовок, название осей, масштаб, описание поля данных и их символическое изображение, если последнее необходимо. Графические символы должны быть четкими, разборчивыми, контрастируя между фоном и рисунком на переднем плане. Каждая из используемых подписей призвана кратко передавать как можно больше информации о том, что график сообщает читателю. При этом подпись не должна содержать краткого изложения или интерпретации результатов, а также деталей эксперимента (Slutsky, 2014).

Важно выбрать правильный тип визуализации в зависимости от характера данных. Круговая и столбчатая диаграммы описывают одномоментную ситуацию, ее результат. Линейный график используют для отображения динамики показателей, он обязательно содержит временную ось. Если автор проводит корреляционный анализ, целесообразно использовать скаттерограмму для демонстрации связи двух показателей. В случае с обоснованием тенденции по показателям обычно применяется гистограмма.

При составлении графика одной из ошибок разработчиков является его усложнение: в одном графике авторы порой объединяют большое количество различных данных. Это затрудняет процесс восприятия материала, читатель может не уловить суть визуализации данных. Определенное затруднение может возникнуть и в случае использования большого количества слов в графике. Его нужно постараться сократить до минимума, при этом следить за тем, чтобы смысл визуального сообщения не потерялся.

Используемые символы должны быть четкими, чтобы не возникло путаницы при их прочтении. Для дополнительной информации по графику можно использовать сноски, при этом удалив всю лишнюю информацию из графика. Занимаясь форматированием графика, необходимо обеспечить его читаемость: учесть тип шрифта, его размер, используемые символы, обратить внимание на выбор цветов и оттенков, название графика, названия осей (Slutsky, 2014).

Соблюдение норм визуализации данных в статье способствует лучшей коммуникации автора с читателями. Благодаря сжатию объемных данных в график научный текст приобретает структурированную форму и становится более доступным для восприятия. Позиция исследователя выглядит при этом более аргументированной и убедительной.

Список литературы

- Bienkowski, M., Feng, M., Means, B. Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief // Office of Educational Technology, US Department of Education. – 2012.
- Glazer, N. Challenges with graph interpretation: A review of the literature // Studies in science education. – 2011. – Vol. 47. – No. 2. – Pp. 183–210. – DOI:10.1080/03057267.2011.605307
- Larose, D. T., Larose, C. D. Discovering knowledge in data: an introduction to data mining. – John Wiley & Sons, 2014.
- Maaten, L., Hinton, G. Visualizing data using t-SNE // Journal of machine learning research. – 2008. – Vol. 9. – Pp. 2579–2605.
- Slutsky, D. J. The effective use of graphs // Journal of wrist surgery. – 2014. – Vol. 3. – No. 02. – Pp. 067–068. – DOI:10.1055/s-0034-1375704
- Yau, N. Visualize this: the FlowingData guide to design, visualization, and statistics. – John Wiley & Sons, 2024. – 384 p.