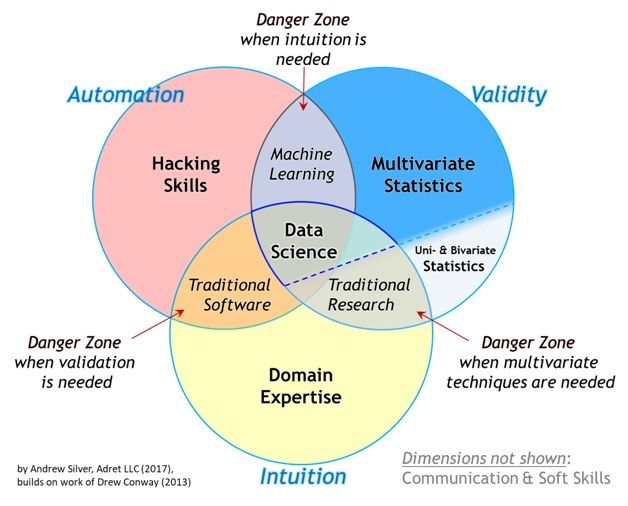
Muslimov Arthur

(since 1.01.20)

*Что такое* ***наука о данных****?*

Долгое время это понятие критиковалось и воспринималось просто как новомодная фраза для украшения резюме работников. Оно и понятно, ведь какая наука не работает с данными? Но критики упустили одну очень важную деталь – Междисциплинарность, ключ к её пониманию. Лучшее объяснение этому дал Венн в своей диаграмме, опубликованной Дрю Конвеем в 2010.



*Почему Python?*

Тут всё просто. Python пригоден для науки о данных в основном благодаря большой и

активно развивающейся экосистеме пакетов, заточенных под научные вычисления.

* библиотеки Numpy – для работы с однородными данными в виде массивов;
* библиотеки Pandas – для работы с неоднородными и поименованными данными;
* SciPy – для общих научных вычислительных задач;
* библиотеки Matplotlib – для визуализации типографического качества;
* оболочки IPython – для интерактивного выполнения и совместного использования кода;
* библиотеки Scikit-Learn – для машинного обучения и множества других инструментов, которые я упомяну в будущем;

Популярным инструментом для разработки на Python служит дистрибутив Anaconda.