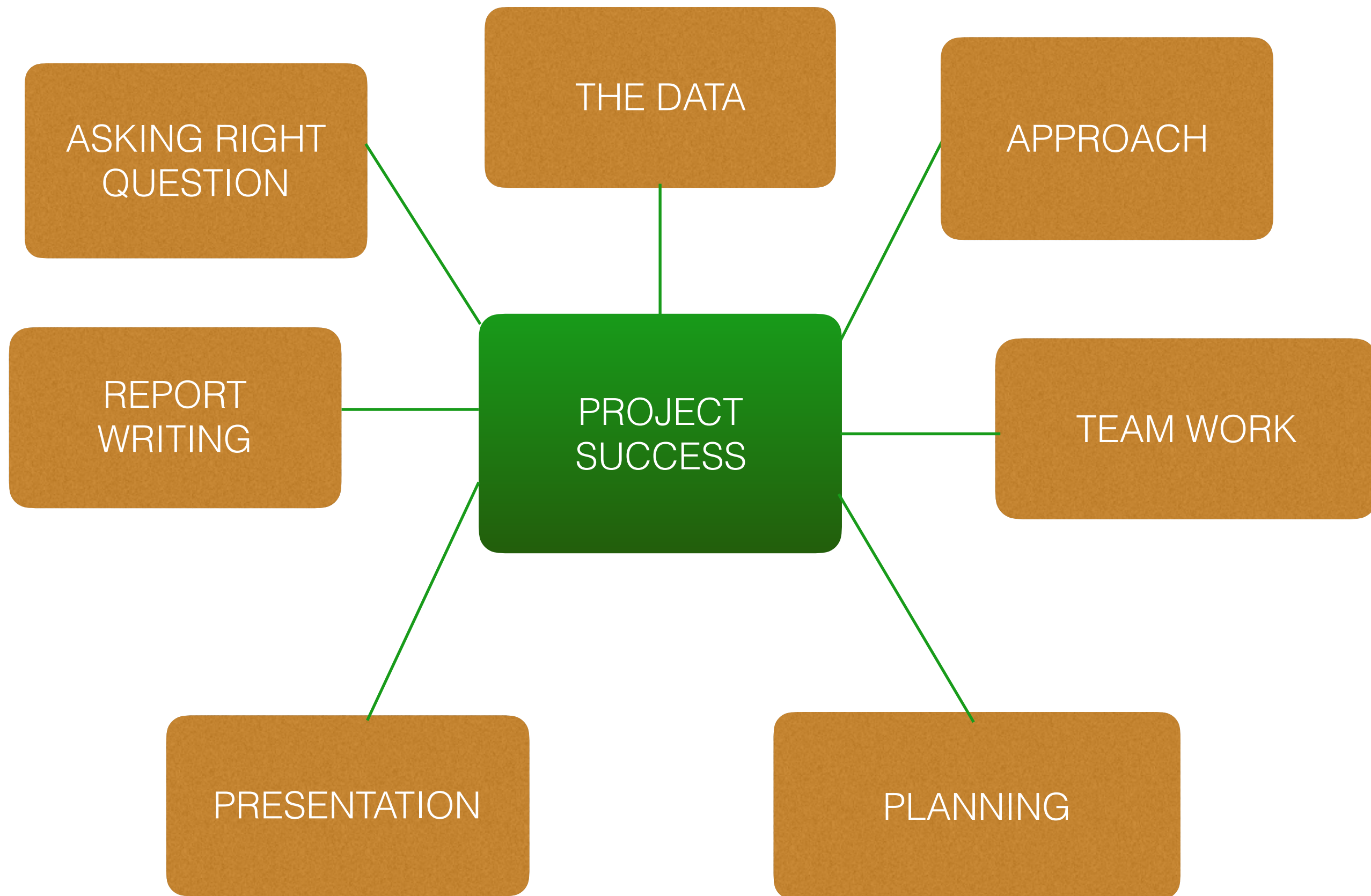


Applied Data Science Fall 2019

Principles of efficient writing

Prof. Stanislav Sobolevsky



Problem statement/research question

One most important point to make clear

- Justification for the study
- Motivation for the audience
- Motivation for the team

Ideally - one phrase

Discussion: formulate the concise problem statement for the following

Consider several horizontal construction projects in NYC. Use MasterCard Retail Location Insight Data to analyze level of commercial activity in different categories of local stores before, after and during the construction. Compare against the nearby areas where construction did not happen.

How could you best formulate the problem statement/research question for the above?

Do horizontal construction projects affect local commercial activity?

Discussion: formulate the concise problem statement for the following-2

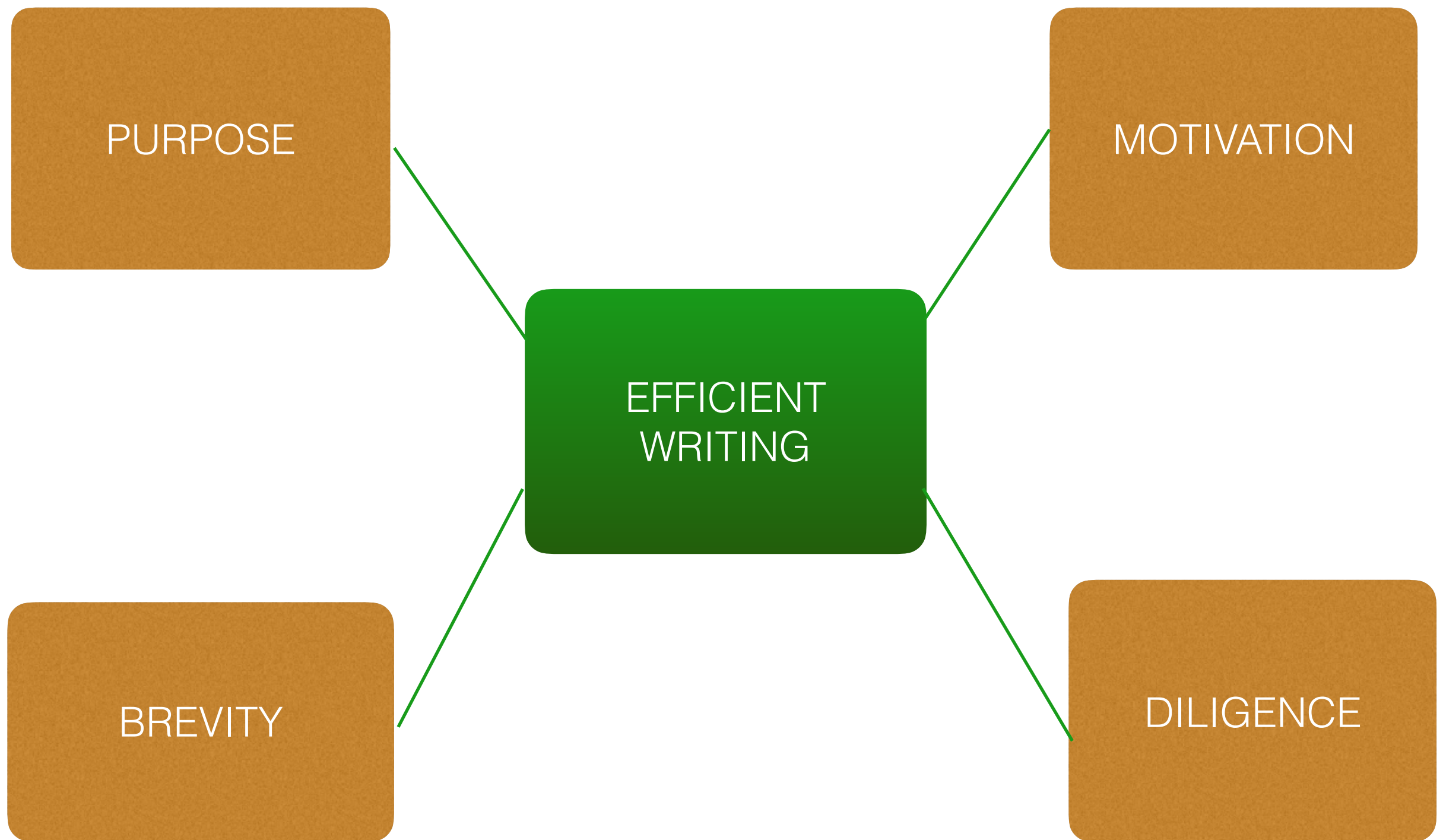
For a sample of small businesses define and quantify various characteristics of customer behavior in the small businesses of interest and nearby competitors based on anonymized credit card transactions. Perform comparative analysis of those characteristics for the businesses who managed to successfully maintain their loans and those who failed. Establish relations between the above characteristics and business default.

How could you best formulate the problem statement/research question for the above?

Can the credit card transaction data help assessing the credit risk of a small business?



Doing research might be hard...
but reading about it should be easy!



Efficient writing should:

- conquer attention
- retain attention as long as needed
- convey the message
- take as little effort from the reader as possible
- ideally - entertain
- never - annoy

Common writing issues and fixes

- Unclear purpose/message
- Redundancy (wordiness, repetitiveness)
- Unsupported claims
- Over/underselling
- Lack of detail/definitions
- Excessive focus on technical details
- Lack of reflection
- Lack of proofreading
- Insufficient time/effort
- Formatting issues
- Clear/concise problem statement, abstract, summary
- Recap in conclusions
- Build a narrative and stick to it
- What min detail reader needs?
- Reflect and iterate
- Cut ruthlessly
- Proofread
- Plan ahead
- Follow formatting requirements/guidelines

Basic formatting expectations

- consistent formatting style (fonts, titles, labels, references, layout etc)
- figures/tables numbered, explained in captions
- all figures/tables/appendices/sources referred to in the text
- consistent referencing style
- consistent bibliography style
- following page/word/figure limits

The language

- Convey clear messages in the most efficient way
- No redundancy
- Consistent style
- Entertaining as much as the style allows, not formal/boring
- Grammar
- Write a draft then edit/polish
- Use proofreading unless proficient

Clear sentences

- Clear message
- Connection with the purpose
- Short sentences (break if necessary)
- Avoid redundant words
- Avoid repeating words

Sentence example - redundancy

Out of 5020 companies **we examined from our** considered sample, only 200 were found risky, **while the remaining 4820 were classified as safe.**

Out of 5020 companies from the considered sample, only 200 were found risky.

Sentence example - length

In our intermittent dosing models, drug resistance took 4-5 months to develop and was inefficient (only ~25% of tumors developed resistance), and so we decided to develop an accelerated and more efficient model of drug resistance by turning to a continuous dosing strategy with 786-O xenografts treated with R32 continuously for 55 days.

In our intermittent dosing models, drug resistance took 4-5 months to develop and was inefficient.

Indeed, only ~25% of tumors developed resistance.

To develop an accelerated and more efficient model of drug resistance, we turned to a continuous dosing strategy.

This way 786-O xenografts were treated with R32 continuously for 55 days.

How to master it - hints

- Writing is not about inborn talent. It is just systematic work
- Do not wait for inspiration - write today!
- Do not try to write perfectly. Make a quick-and-dirty draft. Then revise
- Cut ruthlessly - all words/phrases that don't add value should be gone;
Valuable but less significant stuff - moved to appendixes/Sl.
- Identify confusing/misleading/boring content.
- Take risks. Put something funny or provocative in your writing.

“So what” game

We analyze... We study... We consider... We found that...

Why should the reader (reviewer, editor, sponsor...) care?

For a deeper analysis of temporal variation of customer activity, we look at how time series of daily number of customer transactions can be decomposed into the combination of deterministic and stochastic components.

Why does one ever need this “deeper analysis”? What question does it answer? Purpose?

In order to detect possible anomalies in customer activity, which might point out to the events of interest, one needs to distinguish between the regular temporal patterns subject to usual fluctuations on one hand and unusually strong deviations on the other. For that purpose we consider a decomposition of the time series of daily number of customer transactions into the combination of deterministic and stochastic components.

Writing strategy

1. Problem statement
2. Skeleton, outline of the story
3. Figures and tables. Captions
4. Connect the dots - results
5. Data and Methods
6. Discussion
7. Conclusion
8. Introduction
9. Abstract/executive summary
10. Select Keywords
11. Title
12. Acknowledgements
13. Format references
14. Appendices/Sl

Good start - pitch your paper!

There is only one chance to make first impression!

Title

Abstract

Keywords

may hold the key to publication/business report success

play a pivotal role in the communication of research

- title and abstract/summary are often the only parts freely available or
- read by sponsors, prospective clients, editors
- electronic search engines, databases, or journal websites

Title

Objective: capture attention That's it!

One line. Don't try to tell everything there

Purpose:

- Pitches the content in a few words
- Captures the readers' attention
- Differentiates the paper/report/proposal from others

Should be:

- Catchy
- Relevant
- Keep it short - no redundancy
 - subtitle

Title - recommendations

Keep it simple, brief and attractive

Avoid: On the application of XYZ as a tool for on urban policy-making and decision-making processes

Better: XYZ for urban policy and decision-making

Avoid redundant meaningless words

Avoid: A novel efficient high-performance approach for network community detection

Better: A high-performance approach for network community detection

Specify. Use appropriate descriptive words

Avoid: Impacts of construction projects

Better: Impact of horizontal construction on local commercial activity

Avoid abbreviations and jargon other than commonly used

Discussion - come up with a good title

Come up with a title for the following examples

We analyze the anonymous communication patterns of 2.5 million customers of a Belgian mobile phone operator. Grouping customers by billing address, we build a social network of cities that consists of communications between 571 cities in Belgium. We show that inter-city communication intensity is characterized by a gravity model: the communication intensity between two cities is proportional to the product of their sizes divided by the square of their distance.

In this paper we investigate variations in the adoption of LEED (Leadership in Energy and Environmental Design; one of the features for a “green” building) - certified commercial buildings across 174 core-based statistical areas in the United States. Drawing upon a unique database and using a robust analytical framework, the determinants of the proportion LEED-certified space are modeled. We find that, despite high growth rates, LEED-certified stock accounts for a relatively small proportion of the total commercial stock - the average proportion is less than 1%. Strongest predictors of the proportion of LEED-certified commercial space in a local market are market size, educational attainment and economic growth. In terms of policy effectiveness, it is found that only a mandatory requirement to obtain LEED certification for new buildings has a significant positive effect on market penetration.

Abstract

Should:

- capture attention, motivate to read the paper
- tell a research question
- explain why it is important
- summarize the paper contribution and novelty
- help indexing, contain keywords
- follow word count (usually 150-300 words)

Should not be:

- too long
- redundant
- unfocused
- reporting “whats” without “why’s”
- list all the results (it’s conclusion’s role)
- give definitions, formulas, technical detail etc
- inconsistent with the paper

Abstract - anti-checklist

- new information that is not present in the paper
- undefined abbreviations or group names
- a discussion of previous literature or reference citations
- unnecessary details about the methods use
- big formulas
- figures and tables
- any redundant information or wording

Abstract - checklist

What is the motivation for the research? **MOTIVATION**

What problem is the paper trying to solve? **PROBLEM STATEMENT**

How does the paper achieve the objective? **APPROACH**

What has been found? **RESULTS**

What are the implications of the findings? **CONCLUSIONS**

Abstract example

Introduced in its contemporary form in 1946, but with roots that go back to the eighteenth century, the gravity law is the prevailing framework with which to predict population movement, cargo shipping volume and inter-city phone calls, as well as bilateral trade flows between nations. Despite its widespread use, it relies on adjustable parameters that vary from region to region and suffers from known analytic inconsistencies. Here we introduce a stochastic process capturing local mobility decisions that helps us analytically derive commuting and mobility fluxes that require as input only information on the population distribution. The resulting radiation model predicts mobility patterns in good agreement with mobility and transport patterns observed in a wide range of phenomena, from long-term migration patterns to communication volume between different regions. Given its parameter-free nature, the model can be applied in areas where we lack previous mobility measurements, significantly improving the predictive accuracy of most of the phenomena affected by mobility and transport processes

Filippo Simini et al. *Nature* **484**, 96–100 (05 April 2012) doi:10.1038/nature10856

Abstract example

Introduced in its contemporary form in 1946, but with roots that go back to the eighteenth century, the gravity law is the prevailing framework with which to predict population movement, cargo shipping volume and inter-city phone calls, as well as bilateral trade flows between nations.

Despite its widespread use, it relies on adjustable parameters that vary from region to region and suffers from known analytic inconsistencies.

Here we introduce a stochastic process capturing local mobility decisions that helps us analytically derive commuting and mobility fluxes that require as input only information on the population distribution.

The resulting radiation model predicts mobility patterns in good agreement with mobility and transport patterns observed in a wide range of phenomena, from long-term migration patterns to communication volume between different regions.

Given its parameter-free nature, the model can be applied in areas where we lack previous mobility measurements, significantly improving the predictive accuracy of most of the phenomena affected by mobility and transport processes

MOTIVATION

PROBLEM STATEMENT

APPROACH

RESULTS

CONCLUSIONS

Useful online resources on efficient writing:

<http://writingcenter.unc.edu/handouts>

<http://www.learnhigher.ac.uk/writing-for-university/report-writing/what-makes-a-good-report/>

<https://www.sciencedocs.com/writing-a-scientific-manuscript/>

<http://courses.chem.psu.edu/chem36/Scientific%20Writing.pdf>

<http://www.editage.com/insights/3-basic-tips-on-writing-a-good-research-paper-title>

<https://unilearning.uow.edu.au/report/4bi1.html>

<https://www.proposify.com/blog/executive-summary>

<https://thesistips.wordpress.com/2012/03/25/how-to-write-your-introduction-abstract-and-summary/>

<https://www.elsevier.com/connect/11-steps-to-structuring-a-science-paper-editors-will-take-seriously>

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3626472/>